THE UNIVERSITY OF ALBERTA

PRIMITIVE SEQUENCES AND SEQUENCES OF POSITIVE UPPER LOGARITHMIC DENSITY

by

(C) HARRY R. HENSHAW

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS

EDMONTON, ALBERTA

SPRING, 1972

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "PRIMITIVE SEQUENCES AND SEQUENCES OF POSITIVE UPPER LOGARITHMIC DENSITY", submitted by HARRY R. HENSHAW in partial fulfilment of the requirements for the degree of Master of Science.

ABSTRACT

An increasing sequence, A, of natural numbers is primitive if none of its terms divides any other. Behrend proved that every primitive sequence must have upper asymptotic density less than 1/2 and Erdös showed that the logarithmic density (and therefore lower asymptotic density) of any such sequence must be equal to zero. Besicovitch proved that, for any $\varepsilon > 0$, there exists a primitive sequence with upper density greater than $1/2 - \varepsilon$.

Using a combinatorial result of Sperner, Behrend proved the existence of a positive absolute constant c, such that, for any primitive sequence $A = \{a_i\}$ and any $n \geq 3$

$$\sum_{a_i \leq n} \frac{1}{a_i} < c \frac{\log n}{\sqrt{\log\log n}} .$$

On the other hand, Pillai showed that there is a positive absolute constant c', such that, for any $n \geq 3$, there is a primitive sequence $A' \subseteq \{1,2,\ldots,n\}$, for which

$$\sum_{a \in A'} \frac{1}{a} > c' \frac{\log n}{\sqrt{\log\log n}}$$

Erdös, Sárközi, and Szemerédi showed that

$$\limsup_{n\to\infty} \left\{ \frac{\sqrt{\operatorname{loglog} n}}{\log n} \sum_{a\in A} \frac{1}{a} \right\} = \frac{1}{\sqrt{2\pi}}$$

where the supremum is taken over all primitive sequences $A \subseteq \{1,2,\ldots,n\}$. They also showed that if A is an infinite primitive sequence, then

$$\sum_{a_i \le x} \frac{1}{a_i} = o\left(\frac{\log x}{\sqrt{\operatorname{loglog} x}} \right) \quad \text{as} \quad x \to \infty .$$

Davenport and Erdös showed that every sequence of positive upper logarithmic density has a subsequence, each term of which divides the succeeding term.

Erdös, Sárközi, and Szemerédi obtained the following result. Denoting by $f(x)$ the sum

$$\sum_{\substack{s_i < s_j \le x \\ s_i | s_j}} 1 ,$$

if S has upper logarithmic density $c_1 > 0$, then for infinitely many x,

$$f(x) > x \exp\{c_2 \sqrt{\operatorname{loglog} x}\ \operatorname{logloglog} x\} ,$$

where $c_2 = c_2(c_1)$. There exists, however, a sequence S' having upper logarithmic density c_1, for which

$$f(x) < x \exp\{c_3 \sqrt{\operatorname{loglog} x}\ \operatorname{logloglog} x\}$$

for all x, where $c_3 = c_3(c_1)$.

ACKNOWLEDGEMENTS

I wish to thank Dr. H. L. Abbott for his guidance and assistance in the preparation of this thesis.

I would also like to thank my wife, Marlene, who typed the manuscript.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION AND PRELIMINARY RESULTS FROM NUMBER THEORY AND COMBINATORICS

Let $A = \{a_i\}$ be a finite or infinite sequence of positive integers, $a_1 < a_2 < \ldots$. A is said to be primitive if no a_i divides any other. For example, the integers m, $n < m \leq 2n$, for a fixed n are a finite primitive sequence, and the prime numbers are an infinite primitive sequence.

Suppose $S = \{s_i\}$ is any increasing sequence of positive integers. We will denote the number of $s_i \leq n$ by $S(n)$. The lower (asymptotic) density of S, $\underline{d}S$ is then defined by

$$\underline{d}S = \liminf_{n \to \infty} S(n)/n$$

and the upper density $\bar{d}S$ by

$$\bar{d}S = \limsup_{n \to \infty} S(n)/n \; .$$

If $\underline{d}S = \bar{d}S$, then S is said to have (asymptotic) density dS equal to the common value of the lower and upper densities. The lower logarithmic density of S, $\underline{\delta}S$ is defined by

$$\underline{\delta}S = \liminf_{n \to \infty} \frac{1}{\log n} \sum_{s_i \leq n} \frac{1}{s_i}$$

and the upper logarithmic density and logarithmic density are defined analogously. It can be easily proven (see Halberstam and Roth [10])

that $0 \le \underline{d}S \le \underline{\delta}S \le \bar{\delta}S \le \bar{d}S \le 1$.

We will now briefly outline the contents of this thesis.

Chapter II is devoted to a study of primitive sequences. It can be verified that if A is a primitive sequence, then

$$\bar{d}A \le 1/2 \ . \tag{1.0.1}$$

In fact, (1.0.1) follows from the observation that any two distinct terms of A must have different largest odd divisors. Behrend [2] was the first to prove that equality cannot hold in (1.0.1). On the other hand, Besicovitch [3] proved that, if $\varepsilon > 0$ is given, there exists a primitive sequence A such that $\bar{d}A > 1/2 - \varepsilon$.

Erdös [6] proved that every primitive sequence A has logarithmic density, but not necessarily asymptotic density, by showing that $\bar{\delta}A = 0$, so that $\underline{d}A = \delta A = 0$. This result lead naturally to the consideration of the behaviour of

$$g'(n) = \sum_{a_i \le n} \frac{1}{a_i}$$

for infinite primitive sequences A, and

$$g(n) = \sup_{A} \sum_{a \varepsilon A} \frac{1}{a}$$

where the supremum is taken over all primitive sequences $A \subseteq \{1,\dots,n\}$.

Behrend [2] proved, using a combinatorial result of Sperner,

that for some absolute c,

$$g'(n) \le g(n) < \frac{c}{\sqrt{2\pi}} \frac{\log n}{\sqrt{\log\log n}} .$$

(That $g'(n) = o(\log n)$ is just the result of Erdös, namely $\delta A = 0$).

Pillai [12] proved that if A is the set of square-free integers in $\{1,\ldots,n\}$ having exactly $[\log\log n]$ prime factors, then

$$\sum_{a\varepsilon A} \frac{1}{a} > c_1 \frac{\log n}{\sqrt{\log\log n}} .$$

This lead him to make the following conjecture:

$$L = \lim_{n\to\infty} g(n) \frac{\sqrt{\log\log n}}{\log n} \quad \text{exists.}$$

Erdös [5] claimed that one could modify Behrend's method to show that for every primitive sequence $A \subsetneq \{1,\ldots,n\}$, one has

$$\sum_{a\varepsilon A} \frac{1}{a} \le \left(\frac{1}{\sqrt{2\pi}} + \varepsilon\right) \frac{\log n}{\sqrt{\log\log n}}$$

where $\varepsilon > 0$ is arbitrary and $n > n_0(\varepsilon)$, which is in fact true if A consists of square-free integers. In light of this, he also claimed that Pillai's conjecture was true, and that $L = \frac{1}{\sqrt{2\pi}}$. However, I. Anderson [1] very carefully analyzed the Behrend argument and showed that the best one could hope to get, by this method, is

$$\sum_{a\varepsilon A} \frac{1}{a} < \left(\frac{1}{\sqrt{\pi}} + \varepsilon\right) \frac{\log n}{\sqrt{\log\log n}} .$$

Pillai's conjecture was finally settled in the affirmative, with $L = \frac{1}{\sqrt{2\pi}}$, by Erdös, Sárközi and Szemerédi, but their argument differed substantially from that which Erdös had envisaged in [5].

Erdös, Sárközi, and Szemerédi [8] proved that the analogue of Pillai's theorem does not hold for infinite primitive sequences, by showing that for any such sequence,

$$g'(n) = o\left(\frac{\log n}{\sqrt{\log\log n}}\right).$$

In the same paper, they showed that this result is best possible in the sense that if w(n) is a sequence that tends to zero arbitrarily slowly as n goes to infinity, then there exists an infinite primitive sequence A such that

$$g'(n) > w(n)\frac{\log n}{\sqrt{\log\log n}}$$

for infinitely many n.

In Chapter III of this thesis, we investigate sequences which have positive upper logarithmic density. It follows that such a sequence contains infinitely many pairs of integers a_i, a_j such that $a_i | a_j$. This result can be extended in two ways.

One of these is the famous Davenport-Erdös theorem [4] which asserts that any sequence of positive upper logarithmic density must contain an infinite division chain, i.e., an infinite subsequence, each term of which divides the next.

Suppose that $\overline{\delta}A = c > 0$. Let us denote by $f(x,c)$, the number of pairs $1 \le a_i < a_j \le x$ such that $a_i | a_j$. Then it is clear that $f(x,c) \to \infty$ as $n \to \infty$. Erdös, Sárközi, and Szemerédi obtained some reasonably precise information about the behaviour of $f(x,c)$, and we shall also take up this result in Chapter III.

We have obtained an improved bound in Lemma 2.11 and elaborated some sketched arguments. We have also corrected a number of errors in various papers.

We now give some number theoretic results which we shall use subsequently.

<u>Theorem 1.1</u> Let $a_1 < \ldots < a_k \le n$ be a sequence of positive integers. Let $d^*(m)$ be the number of divisors of m among the a's. Then

$$\sum_{m=1}^{n} d^*(m) = \sum_{i=1}^{k} \left[\frac{n}{a_i}\right].$$

<u>Proof</u>. $\sum_{m=1}^{n} d^*(m)$ is the number of solutions of $xy = m$ for $m \le n$ where x is an a_i and y is an integer. But for each i, the number of solutions of $a_i y \le n$ is $\left[\frac{n}{a_i}\right]$. Thus the total number of solutions is also $\sum_{i=1}^{k}\left[\frac{n}{a_i}\right]$.

We define $d(n)$ to be the number of divisors of n. As

$$\sum_{m=1}^{n} \frac{1}{m} = \log n + O(1) \text{ , we have } \sum_{m=1}^{n} d(m) = n \log n + O(n) .$$

We shall use the following well known results. For p prime

$$\sum_{p \le x} \frac{\log p}{p} = \log x + O(1) \tag{1.1.1}$$

and

$$\sum_{p \le x} \frac{1}{p} = \log\log x + a + o(1) , \tag{1.1.2}$$

where a is a positive constant.

<u>Theorem 1.2</u> (Mertens) $\prod_{p \le x} \left(1 - \frac{1}{p}\right) \sim b \frac{1}{\log x}$, where b is a positive constant.

<u>Theorem 1.3</u> Let $w > 0$ and let $\{t_i\}$ be the sequence of integers which have no prime factors less than w. Then

$$\sum_{t_i \le x} \frac{1}{t_i} > c \frac{\log x}{\log w} , \tag{1.3.1}$$

where c is a positive constant, and x is sufficiently large.

<u>Proof</u>. Let $T(x) = \sum_{t_i \le x} 1$. Let $p_1, p_2, \ldots, p_r$ be the primes not greater than w. Then

$$T(x) = x - \sum_{p_i \le w} \left[\frac{x}{p_i}\right] + \sum_{p_i < p_j \le w} \left[\frac{x}{p_i p_j}\right] - \cdots$$

$$= x - \sum_{p_j \le w} \frac{x}{p_i} + \sum_{p_i < p_j \le w} \frac{x}{p_i p_j} - \dots + 0(2^r) ,$$

since there are 2^r summands. Thus

$$T(x) = x \prod_{p \le w} \left(1 - \frac{1}{p}\right) + 0(2^r)$$

$$\sim b \frac{x}{\log w} + 0(2^r) ,$$

by Mertens' Theorem.

Then, by partial summation,

$$\sum_{t_i \le x} \frac{1}{t_i} = T(x) \cdot \frac{1}{x} + \int_w^x T(t) \frac{1}{t^2} dt$$

$$\sim b \frac{1}{\log w} + 0\left(2^r \cdot \frac{1}{x}\right) + b \frac{1}{\log w} \int_w^x \frac{dt}{t}$$

$$+ 0\left(2^r \int_w^x \frac{dt}{t^2}\right)$$

$$\sim b \frac{\log x}{\log w} - b + b \frac{1}{\log w} + 0\left(2^r \cdot \frac{1}{x}\right)$$

$$+ 0\left(2^r\left(\frac{1}{w} - \frac{1}{x}\right)\right)$$

$$> \frac{b}{2} \frac{\log x}{\log w} ,$$

if x is sufficiently large, say $x > \exp \exp w$.

We denote by $\omega(m)$ the number of different primes dividing m and by $\Omega(m)$ the number of prime powers dividing m. Similarly we will denote by $\omega_T(m)$ the number of different primes less than T which divide m and by $\Omega_T(m)$ the number of prime powers, for primes less than T, which divide m.

<u>Theorem 1.4</u> For $T < n$, $\sum_{m=1}^{n} \{\Omega_T(m) - A_n\}^2 = O(n \operatorname{loglog} T)$,

where $A_n = \sum_{p \le n} \frac{\omega_T(p)}{p}$.

<u>Proof</u>. We will find it more convenient to work with $\omega_T(m)$ so we first replace $\Omega_T(m)$ by $\omega_T(m)$. We have

$$\Omega_T(m) - \omega_T(m) = \sum_{p} \omega_T(p) \sum_{\substack{k=2 \\ p^k | m}}^{\infty} 1 .$$

We show

$$\sum_{m=1}^{n} \{\Omega_T(m) - \omega_T(m)\}^2 = O(n) . \tag{1.4.1}$$

Now $\sum_{m=1}^{n} \{\Omega_T(m) - \omega_T(m)\}^2 = \sum_{p} \sum_{q} \omega_T(p)\omega_T(q) \sum_{j=2}^{\infty} \sum_{k=2}^{\infty} \left(\sum_{\substack{m=1 \\ p^j | m, q^k | m}}^{n} 1 \right)$

$$= \sum_{\substack{p \quad q \\ p \neq q}} \sum \omega_T(p)\omega_T(q) \sum_{j=2}^{\infty} \sum_{k=2}^{\infty} \left[\frac{n}{p^j q^k} \right]$$

$$+ \sum_{p} \omega_T^2(p) \sum_{j=2}^{\infty} \sum_{k=2}^{\infty} \left[\frac{n}{p^{\max\{j,k\}}} \right] .$$

We have $\displaystyle\sum_{p} \sum_{\substack{q \\ p \neq q}} \omega_T(p)\omega_T(q) \sum_{j=2}^{\infty} \sum_{k=2}^{\infty} \left[\frac{n}{p^j q^k}\right] = O\left(n \sum_{p} \sum_{\substack{q \\ p \neq q}} \omega_T(p)\omega_T(q) \frac{1}{p^2 q^2} \right)$

$$= O\left(n \left(\sum_{p} \frac{\omega_T(p)}{p^2} \right)^2 \right) = O(n) ,$$

and $\displaystyle\sum_{p} \omega_T(p) \sum_{j=2}^{\infty} \sum_{k=2}^{\infty} \left[\frac{n}{p^{\max\{j,k\}}} \right] = O\left(n \sum_{p} \omega_T^2(p) \sum_{k=2}^{\infty} \frac{k}{p^k} \right)$

$$= O\left(n \sum_{p} \frac{\omega_T^2(p)}{p^2} \right) = O(n) ,$$

(since $\displaystyle\sum_{k=2}^{\infty} \frac{k}{p^k} \le \sum_{k=2}^{\infty} \left(\frac{2}{p}\right)^k = O\left(\frac{1}{p^2}\right)$) which proves (1.4.1) .

For all a, b, $(a+b)^2 \le 2a^2 + 2b^2$, so

$$\sum_{m=1}^{n} \{\Omega_T(m) - A_n\}^2 \le 2 \sum_{m=1}^{n} \{\Omega_T(m) - \omega_T(m)\}^2$$

$$+ 2 \sum_{m=1}^{n} \{\omega_T(m) - A_n\}^2$$

and we have only to show that

$$\sum_{m=1}^{n} \{\omega_T(m) - A_n\}^2 = O(n \log\log T) . \tag{1.4.2}$$

Now

$$\sum_{m=1}^{n} \{\omega_T(m) - A_n\}^2 = \sum_{m=1}^{n} \omega_T^2(m) - 2A_n \sum_{m=1}^{n} \omega_T(m) + nA_n^2 \tag{1.4.3}$$

and

$$\sum_{m=1}^{n} \omega_T(m) = \sum_{p} \omega_T(p) \sum_{\substack{m=1 \\ p|m}}^{n} 1$$

$$= \sum_{p} \omega_T(p) \left[\frac{n}{p}\right] = nA_n + O(n) . \tag{1.4.4}$$

Also

$$\sum_{m=1}^{n} \omega_T^2(m) = \sum_{\substack{p \\ p \neq q}} \sum_{q} \omega_T(p)\omega_T(q) \left[\frac{n}{pq}\right] + \sum_{p} \omega_T^2(p) \left[\frac{n}{p}\right]$$

$$= n \sum_{\substack{pq \leq n \\ p \neq q}} \frac{\omega_T(p)\ \omega_T(q)}{pq} + O\left(\sum_{pq \leq n} 1\right) + nA_n + O(n)$$

$$= n \sum_{pq \leq n} \frac{\omega_T(p)\ \omega_T(q)}{pq} + nA_n + O(n)$$

since every m has at most two representations $m = pq$. We have

$$A_{\sqrt{n}}^2 = \left(\sum_{p \leq \sqrt{n}} \frac{\omega_T(p)}{p}\right)^2 \leq \sum_{pq \leq n} \frac{\omega_T(p)\ \omega_T(q)}{pq} \leq \left(\sum_{p \leq n} \frac{\omega_T(p)}{p}\right)^2 = A_n^2 .$$

As
$$A_n - A_{\sqrt{n}} = \sum_{\sqrt{n}<p\le n} \frac{\omega_T(p)}{p} \le \sum_{\sqrt{n}<p\le n} \frac{1}{p} = O(1) ,$$

$$A_n^2 - A_{\sqrt{n}}^2 = O(A_n + A_{\sqrt{n}}) = O(A_n) .$$

Hence

(1.4.5)
$$\sum_{m=1}^{n} \omega_T^2(m) = n A_n^2 + O(n\{A_n + 1\}) .$$

Therefore, by (1.4.3), (1.4.4), and (1.4.5),

$$\sum_{m=1}^{n} \{\omega_T(m) - A_n\}^2 = n A_n^2 + O(n\{A_n + 1\})$$

$$- 2 A_n\{n A_n + O(n)\} + n A_n^2$$

$$= O(n\{A_n + 1\}) .$$

Finally, $A_n = \sum_{p\le n} \frac{\omega_T(p)}{p} = \text{loglog } T + O(1)$, so $n\{A_n + 1\} = O(n \text{ loglog } T)$, which proves the theorem.

Corollary 1.5 Let $\varepsilon > 0$ be given. Then there is a $T_o = T_o(\varepsilon)$ so that if $T_o < T < n$, then the number of positive integers $m \le n$ for which $| \Omega_T(m) - \text{loglog } T | \ge \frac{1}{3} \text{loglog } T$ is less than εn .

Proof. It follows from Theorem 1.4 that if we denote by N_n, the number of integers not exceeding n for which

$$| \Omega_T(m) - \text{loglog } T | \geq \frac{1}{3} \text{loglog } T ,$$

then $$N_n < C \frac{1}{3}^{-2} n(\text{loglog } T)^{-1} .$$

Choose T_0 large enough that $9C (\text{loglog } T)^{-1} < \varepsilon$, and the result follows.

Lastly, we shall find Stirling's formula

$$\log n! = (n + 1/2) \log n - n + 0(1)$$

a useful tool in estimating bounds, particularly for binomial coefficients.

If $S = \{s_1,s_2,\ldots\}$ is a set and $P = \{p_1,p_2\ldots\}$ is a set of primes, then we can make a correspondence by associating s_i with p_i. This induces a correspondence between the non-empty subsets of S and the square-free numbers that can be formed from primes in P. Moreover, set inclusion for subsets of S corresponds to divisibility, that is, $p_{i_1}\ldots p_{i_m} | p_{j_1}\ldots p_{j_n}$ if and only if $\{s_{i_1},\ldots,s_{i_m}\} \subseteq \{s_{j_1},\ldots,s_{j_n}\}$. Thus it should not be surprising that the proofs of many of the theorems on primitive sequences make use of certain combinatorial results, which we now present.

We shall denote the order of a set S by $|S|$. If S is any set, and A any family of subsets of S, we shall denote by $A^{(h)}$ the subfamily of all h-element members of A, and by $B(A)$ the family of all subsets of S which contain members of A. We shall call the family A <u>primitive</u> if none of its members contains any other.

Lemma 1.6 Let S by a set of order n. Let H_j be a family of j-element subsets of S. Let $k > j$ and suppose T_k is any collection of k-element subsets of S, including all those containing a member of H_j. Then

$$\frac{|T_k|}{\binom{n}{k}} \geq \frac{|H_j|}{\binom{n}{j}} \quad .$$

The same result holds if $k < j$ and T_k includes all k-element subsets of S contained in members of H_j.

Proof. We prove the lemma for $k > j$, the other case being similar. The members of H_j can be contained in members of T_k in at least $\binom{n-j}{k-j} |H_j|$ ways and at most $\binom{k}{j} |T_k|$ ways. Hence

$$\binom{n-j}{k-j} |H_j| \leq \binom{k}{j} |T_k| \quad .$$

Since
$$\binom{n-j}{k-j} \binom{k}{j}^{-1} = \binom{n}{k} \binom{n}{j}^{-1}$$

the result follows.

Theorem 1.7 (Sperner [13]) Let S be a set of order n and let A be a primitive family of subsets of S. Then

$$|A| \leq \binom{n}{[n/2]} .$$

Proof. We shall apply Lemma 1.6 repeatedly to replace the family A by a primitive family at least as large, all of whose members have $[n/2]$ elements.

Suppose that the largest members of A have m elements, where $m > [n/2]$. Let us denote by $C^{(m-1)}$ the family of all (m-1)-element subsets of S which are contained in members of $A^{(m)}$. By Lemma 1.6, $|C^{(m-1)}| \geq |A^{(m)}|$. Furthermore, since A is primitive it is apparent that $A \sim A^{(m)} \bigcup C^{(m-1)}$ is also a primitive family whose largest members are of size (m-1). Since A and $C^{(m-1)}$ are disjoint,

$$|A \sim A^{(m)} \bigcup C^{(m-1)}| \geq |A| .$$

After repeating this procedure a sufficient number of times, we obtain a primitive family A_1, each of whose members has at most $[n/2]$ elements, such that

$$|A| \leq |A_1| .$$

Applying a similar procedure to the smallest members of A_1 we will eventually obtain a primitive family A_2, each of whose members has exactly $[n/2]$ elements, such that

$$|A_1| \leq |A_2| .$$

As there are $\binom{n}{[n/2]}$ subsets of S of size $[n/2]$, we therefore have

$$|A| \leq |A_1| \leq |A_2| \leq \binom{n}{[n/2]} .$$

<u>Theorem 1.8</u> (Kleitman [11]) Let S be a set of order n and let A be a primitive family of subsets of S. Then if A has at least $\binom{n}{k}$ members, $k \leq [n/2]$, $B(A)$ has at least $\sum_{j=0}^{k} \binom{n}{j}$ members.

Proof. Our proof will be similar in method to that of Sperner's Theorem. Applying Lemma 1.6 repeatedly we will transform the family A to the family A' of all $(n - k)$-element subsets of S in such a way that

$$|B(A)| \geq |B(A')| = \sum_{j=n-k}^{k} \binom{n}{j} = \sum_{j=0}^{k} \binom{n}{j} .$$

Let $m = \min\{h : A^{(h)} \neq \phi\}$. Without loss of generality, we may assume that $m < n - k$. Let $C^{(m+1)}$ be the maximal family of $(m + 1)$-element subsets of S each of which contains some member of $A^{(m)}$. Since A is primitive, we must have $A \cap C^{(m+1)} = \phi$. If $|C^{(m+1)}| > |A^{(m)}|$, we will denote by $\bar{C}^{(m+1)}$ any $|A^{(m)}|$ - sized subfamily of $A^{(m)}$; otherwise we will let $\bar{C}^{(m+1)}$ be $C^{(m+1)}$.

We define an operation on primitive families by

$$X(A) = A \sim A^{(m)} \cup \bar{C}^{(m+1)} .$$

Then

(1.8.1) $$|X(A)| - |A| = |\bar{C}^{(m+1)}| - |A^{(m)}| .$$

Since $\bar{C}^{(m+1)} \subseteq B(A)$, $B(X(A)) = B(A) \sim A^{(m)}$ and

(1.8.2) $$|B(X(A))| - |B(A)| = - |A^{(m)}| .$$

If $|C^{(m+1)}| \geq |A^{(m)}|$, then

$$|B(X(A))| - |B(A)| < 0 = |X(A)| - |A| .$$

On the other hand, if $|C^{(m+1)}| < |A^{(m)}|$, by Lemma 1.6, we must

have $m > n/2$

and
$$|C^{(m+1)}| \geq |A^{(m)}| \frac{\binom{n}{m+1}}{\binom{n}{m}} .$$

Hence
$$|C^{(m+1)}| - |A^{(m)}| \geq - |A^{(m)}| \left(1 - \frac{\binom{n}{m+1}}{\binom{n}{m}}\right) .$$

Since $m > n/2$, $1 - \frac{\binom{n}{m+1}}{\binom{n}{m}} > 0$. Thus, by (1.8.1) and (1.8.2) we have

(1.8.3)
$$|\mathcal{B}(X(A))| - |\mathcal{B}(A)| \leq K\{|X(A)| - |A|\}$$

in either case, where K is a sufficiently large constant.

As $X(A)$ is primitive we can apply the operation X again. After $n - k - m$ such applications of X, we obtain the primitive family $X^{n-k-m}(A)$, whose smallest members have $n - k$ elements. At each step (1.8.3) holds, so for $i = 1,\ldots,n - k - m$,

$$|\mathcal{B}(X^i(A))| - |\mathcal{B}(X^{i-1}(A))| \leq K\{|X^i(A)| - |X^{i-1}(A)|\} .$$

Then
$$\sum_{i=1}^{n-k-m} \{|\mathcal{B}(X^i(A)| - |\mathcal{B}(X^{i-1}(A))|\} \leq K \sum_{i=1}^{n-k-m} \{|X^i(A)| - |X^{i-1}(A)|\}$$

and

(1.8.4)
$$|\mathcal{B}^{n-k-m}(A))| - |\mathcal{B}(A)| \leq K\{|X^{n-k-m}(A)| - |X(A)|\} .$$

Before we can remove the members of $X^{n-k-m}(A)$ having more than $n - k$ elements, we must ensure that we shall be able to determine how many members of $\mathcal{B}(X^{n-k-m}(A))$ will be removed at the

same time. To do this we modify $X^{n-k-m}(A)$ slightly.

Let us denote by $\mathcal{S}^{(j)}$ the family of j-element subsets of S and let $\mathcal{P}$ be any primitive family of subsets of S. Let

$$p_1 = \max\{h: \mathcal{P}^{(h)} \neq \phi\},$$

$$p_2 = \max\{j: \mathcal{S}^{(j)} \cap \mathcal{B}(\mathcal{P}) \neq \mathcal{S}^{(j)}\},$$

and

$$p = \max\{p_1, p_2\}.$$

We now replace $\mathcal{P}$ by

$$Y(\mathcal{P}) = \mathcal{P} \cup (\mathcal{S}^{(p)} \sim (\mathcal{B}(\mathcal{P}) \cap \mathcal{S}^{(p)})).$$

If $p > p_2$, $Y(\mathcal{P}) = \mathcal{P}$ and $\mathcal{B}(Y(\mathcal{P})) = \mathcal{B}(\mathcal{P})$. Otherwise $\mathcal{B}(Y(\mathcal{P})) = \mathcal{B}(\mathcal{P}) \cup (\mathcal{S}^{(p)} \sim (\mathcal{B}(\mathcal{P}) \cap \mathcal{S}^{(p)}))$, since for all $r > p$ we have $\mathcal{S}^{(r)} \subseteq \mathcal{B}(\mathcal{P})$. Therefore, in either case, $|\mathcal{B}(Y(\mathcal{P}))| - |\mathcal{B}(\mathcal{P})| = |Y(\mathcal{P})| - |\mathcal{P}|$. In particular,

(1.8.5) $|\mathcal{B}(YX^{n-k-m}(A))| - |\mathcal{B}(X^{n-k-m}(A))| = |YX^{n-k-m}(A)| - |X^{n-k-m}(A)|$

Again, let $\mathcal{P}$ be any primitive family of subsets of S and suppose that the largest members of $\mathcal{P}$ have order $p > \left[\frac{n+1}{2}\right]$. If $\mathcal{B}(\mathcal{P}) \supseteq \mathcal{S}^{(r)}$ for $r \geq p$, we shall say that $\mathcal{B}(\mathcal{P})$ is <u>full</u>. Let $\mathcal{D}^{(p-1)} = \mathcal{S}^{(p-1)} \sim (\mathcal{B}(\mathcal{P}) \cap \mathcal{S}^{(p-1)})$. Then $\mathcal{D}^{(p-1)} \cap \mathcal{P} = \phi$.

We define an operation Z on $\mathcal{P}$ by

$$Z(\mathcal{P}) = \mathcal{P} \sim \mathcal{P}^{(p)} \cup \mathcal{D}^{(p-1)}.$$

We then have

(1.8.6) $$|Z(\mathcal{P})| - |\mathcal{P}| = |\mathcal{D}^{(p-1)}| - |\mathcal{P}^{(p)}|.$$

The largest members of $Z(\mathcal{P})$ have $p-1$ elements so $\mathcal{B}(Z(\mathcal{P}))$ is full. If $\mathcal{B}(\mathcal{P})$ is also full, then

(1.8.7) $$|\mathcal{B}(Z(\mathcal{P}))| - |\mathcal{B}(\mathcal{P})| = |\mathcal{D}^{(p-1)}| .$$

Since $\mathcal{P}$ is primitive, every $(p-1)$-element subset of each member of $\mathcal{P}^{(p)}$ is a member of $\mathcal{S}^{(p-1)} \sim (\mathcal{B}(\mathcal{P}) \cap \mathcal{S}^{(p-1)}) = \mathcal{D}^{(p-1)}$. Thus, by Lemma 1.6,

$$|\mathcal{P}^{(p)}| \leq |\mathcal{D}^{(p-1)}| \frac{\binom{n}{p}}{\binom{n}{p-1}} ,$$

and $$|\mathcal{D}^{(p-1)}| - |\mathcal{P}^{(p)}| \geq |\mathcal{D}^{(p-1)}| \left(1 - \frac{\binom{n}{p}}{\binom{n}{p-1}}\right) .$$

Since $p > \left[\frac{n+1}{2}\right]$, $1 - \frac{\binom{n}{p}}{\binom{n}{p-1}} > 0$. This together with (1.8.6) and (1.8.7) gives

(1.8.8) $$|\mathcal{B}(Z(\mathcal{P}))| - |\mathcal{B}(\mathcal{P})| \leq K\{|Z(\mathcal{P})| - |\mathcal{P}|\} ,$$

where we can choose K to be the constant of (1.8.3). Finally, since $\mathcal{D}^{(p-1)} \cap \mathcal{B}(\mathcal{P}) = \phi$, $Z(\mathcal{P})$ is primitive.

The smallest members of the primitive family $YX^{n-k-m}(\mathcal{A})$ have order not less than $n-k \geq \left[\frac{n+1}{2}\right]$, and without loss of generality, the largest members have order $p > n-k$. We can therefore apply Z to $YX^{n-k-m}(\mathcal{A})$ $p-n+k$ times. As $\mathcal{B}(YX^{n-k+m}(\mathcal{A}))$ is full, we have, for $i = 1,\dots, p-n+k$,

$$|\mathcal{B}(Z^iYX^{n-k-m}(A))| - |\mathcal{B}(Z^{i-1}YX^{n-k-m}(A))|$$

$$\leq K\{|Z^iYX^{n-k-m}(A)| - |Z^{i-1}YX^{n-k-m}(A)|\} .$$

Summing over i gives

(1.8.9) $$|\mathcal{B}(Z^{p-n+k}YX^{n-k-m}(A))| - |\mathcal{B}(YX^{n-k-m}(A))|$$

$$\leq K\{|Z^{p-n+k}YX^{n-k-m}(A)| - |YX^{n-k-m}(A))| .$$

Adding the inequalities (1.8.4), (1.8.5) and (1.8.9) gives

$$|\mathcal{B}(Z^{p-n+k}YX^{n-k-m}(A))| - |\mathcal{B}(A)| \leq K\{|Z^{p-n+k}YX^{n-k-m}(A)| - |A|\} .$$

But every member of $Z^{p-n+k}YX^{n-k-m}(A)$ has order $n - k$ and $\mathcal{B}(Z^{p-n+k}YX^{n-k-m}(A))$ is full. Thus $Z^{p-n+k}YX^{n-k-m}(A) = \mathcal{S}^{(n-k)}$ and $\mathcal{B}(Z^{p-n+k}YX^{n-k-m}(A)) = \bigcup_{j=n-k}^{n} \mathcal{S}^{(j)}$. By hypothesis, $|A| \geq \binom{n}{k}$.

Therefore

$$\sum_{j=0}^{k}\binom{n}{j} - |\mathcal{B}(A)| \leq K\{\binom{n}{k} - |A|\} \leq 0$$

and
$$|\mathcal{B}(A)| \geq \sum_{j=0}^{k}\binom{n}{j}$$

proving the theorem.

We will use the theorems of Sperner and Kleitman to prove the following theorem of Erdös, Sárközi, and Szemerédi [8] .

<u>Theorem 1.9</u> Let S, S_1 and S_2 be sets with $S = S_1 \cup S_2$, $S_1 \cap S_2 = \phi$, where $|S_1| = k$, $|S_2| = m$, $m \geq k$. Let $A_1, A_2, \ldots, A_r$

be a primitive family of subsets of S_1, where

$$r > c_1 \frac{2^{k+m}}{\sqrt{k+m}} .$$

Let $B_1, B_2, \ldots, B_t$ be the (distinct) subsets of S of the form

$$A_i \cup R \qquad i = 1, \ldots, r , \tag{1.9.1}$$

with R running through all the subsets of S_2. Then for $m > m_0(c_1)$,

$$t > c_2\ 2^{k+m} ,$$

where $c_2 = c_2(c_1)$.

Proof. We partition the family of A's into 2^k classes, where A_i and A_j are in the same class if and only if $A_i \cap S_1 = A_j \cap S_1$.

If A_i and A_j are in the same class, then $A_i \cap S_2$ cannot contain $A_j \cap S_2$, otherwise A_i would contain A_j . Thus the family $\{A_i \cap S_2\}$, for all A_i in a fixed class, is primitive. By Sperner's theorem and Stirling's formula, each class therefore contains at most

$$\binom{m}{[m/2]} < c \frac{2^m}{\sqrt{m}}$$

A's.

There exists a positive constant $c_3 = c_3(c_1)$ such that at least $c_3 2^k$ classes have more than $c_3 \frac{2^m}{\sqrt{m}}$ members each. For if

not, the total number of A's would be at most

$$c_3 2^k \, c \, \frac{2^m}{\sqrt{m}} + (1 - c_3) \, 2^k \, c_3 \, \frac{2^m}{\sqrt{m}} \le c_3 (c + 1) \, \frac{2^{k+m}}{\sqrt{m}}$$

but at least $$c_1 \frac{2^{k+m}}{\sqrt{k+m}} \ge \frac{c_1}{\sqrt{2}} \, \frac{2^{k+m}}{\sqrt{m}} ,$$

a contradiction for c_3 sufficiently small. Using Kleitman's theorem, we count the number of sets B formed from the A's in one such class $\mathcal{C}$.

Firstly we show that there is a positive constant $c_4 = c_4(c_3) = c_4(c_1)$ such that

$$\binom{m}{p} < c_3 \frac{2^m}{\sqrt{m}} \tag{1.9.2}$$

where $p = \left[\frac{m}{2} - c_4 \sqrt{m} \right]$. By Stirling's formula we have

$$\log \binom{m}{p} = \left(m + \frac{1}{2}\right) \log m - \left(p + \frac{1}{2}\right) \log p$$

$$- \left(m - p + \frac{1}{2}\right) \log (m - p) + 0(1) .$$

$$\le \left(m + \frac{1}{2}\right) \log m - \left(\frac{m}{2} - c_4\sqrt{m} + \frac{1}{2}\right) \log \left(\frac{m}{2} \left(1 - \frac{2c_4}{\sqrt{m}}\right)\right)$$

$$- \left(\frac{m}{2} + c_4\sqrt{m} + \frac{1}{2}\right) \log \left(\frac{m}{2} \left(1 + \frac{2c_4}{\sqrt{m}}\right)\right) + 0(1) .$$

$$= m \log 2 - \frac{1}{2} \log m - \left(\frac{m}{2} + \frac{1}{2}\right) \left(\log\left(1 + \frac{2c_4}{\sqrt{m}}\right) + \log\left(1 - \frac{2c_4}{\sqrt{m}}\right)\right)$$

$$- c_4\sqrt{m} \left(\log\left(1 + \frac{2c_4}{\sqrt{m}}\right) - \log\left(1 - \frac{2c_4}{\sqrt{m}}\right) \right) + 0(1) .$$

Using Taylor series expansions, we have, for m sufficiently large

$$\log\left(1+\frac{2c_4}{\sqrt{m}}\right)+\log\left(1-\frac{2c_4}{\sqrt{m}}\right) = -\frac{4c_4^{\,2}}{m} - 2\int_0^{\frac{2c_4}{\sqrt{m}}} \frac{t^3}{1-t^2}\,dt$$

$$= -\frac{4c_4^{\,2}}{m} + o\left(\frac{1}{m}\right)$$

and

$$\log\left(1+\frac{2c_4}{\sqrt{m}}\right)-\log\left(1-\frac{2c_4}{\sqrt{m}}\right) = \frac{4c_4}{\sqrt{m}} + 2\int_0^{\frac{2c_4}{\sqrt{m}}} \frac{t^2}{1-t^2}\,dt\,,$$

$$= \frac{4c_4}{\sqrt{m}} + o\left(\frac{1}{\sqrt{m}}\right)\ .$$

Therefore

$$\log\binom{m}{p} \leq m\log 2 - \frac{1}{2}\log m + 2c_4^{\,2} - 4c_4^{\,2} + O(1)$$

$$= m\log 2 - \frac{1}{2}\log m - 2c_4^{\,2} + O(1)$$

and after a suitable choice of c_4, we have (1.9.2) .

Similarly, there exists a positive constant $c_5 = c_5(c_3) = c_5(c_1)$, with $c_5 > c_4$, such that

(1.9.3) $$\binom{m}{q} > \frac{c_3}{2}\,\frac{2^m}{\sqrt{m}}\,,$$

where $$q = \left[\frac{m}{2} - c_5\sqrt{m}\right]\ .$$

The family of all sets B of the form (1.9.1) is $\mathcal{B}(\{A_i \cap S_2 : A_i \in \mathcal{C}\})$. By (1.9.2), $|\{A_i \cap S_2 : A_i \in \mathcal{C}\}| = |\mathcal{C}| \geq \binom{m}{p}$.

Hence by Kleitman's theorem, and (1.9.3)

$$|\mathcal{B}(\{A_i \cap S_2 : A_i \in C\})| \geq \sum_{j=0}^{p} \binom{m}{j} \geq \sum_{j=q}^{p} \binom{m}{j}$$

$$> (c_5 - c_4)\, \frac{c_3}{2}\, 2^m = c_6 2^m$$

proving the theorem, since there are $c_3 2^k$ classes with $c_6 2^m$ members.

<u>Theorem 1.10</u> Let S be a set of n elements and let $B_1, \ldots, B_z$, $z > c_1 2^n$ be subsets of S. Then if $n > n_o(c_1)$, one of the B's contains at least $\exp\{c_2\sqrt{n}\log n\}$ of the B's, where $c_2 = c_2(c_1)$.

<u>Proof</u>. We first show that it is sufficient to consider only certain of the B's.

Using Stirling's formula, and an argument similar to that employed in the proof of the previous theorem, it can be seen that for $n > n_o(c_1)$ there exists a constant $c_3 = c_3(c_1)$ such that

$$\sum_{j \leq \frac{n}{2}+c_3\sqrt{n}} \binom{n}{j} + \sum_{j \geq \frac{n}{2}+c_3\sqrt{n}} \binom{n}{j} < \frac{1}{4} c_1 2^n < \frac{1}{4} z .$$

Thus we can assume without loss of generality, that

$$\text{(1.10.1)} \qquad \frac{n}{2} - c_3\sqrt{n} < |B_i| < \frac{n}{2} + c_3\sqrt{n} \qquad i = 1, \ldots, z .$$

Let $\mathcal{B}^{(j)}$ be the family of j-element B's and $g(j) = |\mathcal{B}^{(j)}|$. Since for each j, $g(j) \leq \binom{n}{j} < c\frac{2^n}{\sqrt{n}}$, there must be a positive constant

$c_4 = c_4(c_1)$ such that there are more than $c_4\sqrt{n}$ j's for which $g(j)$ is non-zero. For if not, $z = \sum_j g(j) \le c_4\sqrt{n}\, c\frac{2^n}{\sqrt{n}} < z$, for c_4 sufficiently small.

Denoting by Σ' the summation over all j's for which $g(j) < \frac{1}{2}c_1\binom{n}{j}$ we have

$$\sum_j{}' g(j) \le \frac{1}{2}c_1 \sum_{j=1}^{n}\binom{n}{j} < \frac{1}{2}z \quad .$$

Therefore, we can also assume that either $g(j) = 0$ or $g(j) > \frac{1}{2}c_1\binom{n}{j}$.

Let $$s = \left[\frac{2}{c_1}\right] + 2 \ .$$

For $n > n_0(s) = n_0(c_1)$ there is a sequence $j_1 < \ldots < j_s$ satisfying

(1.10.2) $$g(j_r) > \frac{1}{2}c_1\binom{n}{j_r} \qquad r = 1,\ldots,s$$

since there are more than $c_4\sqrt{n}$ j's for which $g(j) > 0$. Furthermore since there are at most $2c_3\sqrt{n}$ j's, the sequence can be chosen to satisfy

(1.10.3) $$2c_3\sqrt{n} > j_r - j_{r-1} > c_5\sqrt{n} \qquad r = 2,\ldots,s$$

where $c_5 = c_5(c_1) > c_3$.

By Stirling's formula, (1.10.1) and (1.10.2) we have

$$\log\binom{j_r}{j_{r-1}} = j_r \log j_r - \left(j_{r-1} + \frac{1}{2}\right)\log j_{r-1}$$

$$- \left(j_r - j_{r-1} + \frac{1}{2}\right) \log j_r - j_{r-1} + 0(1)$$

$$> (j_r - j_{r-1}) \log j_r - \left(j_r - j_{r-1} + \frac{1}{2}\right) \log(j_r - j_{r-1}) + 0(1)$$

$$> c_5\sqrt{n} \log n - 2c_3\sqrt{n} \log \sqrt{n} + o(\sqrt{n} \log n)$$

$$> c_6\sqrt{n} \log n .$$

Thus

(1.10.4) $$\binom{j_r}{j_{r-1}} > \exp\{c_6\sqrt{n} \log n\} .$$

Denoting by $B_1^*, \ldots, B_t^*$ all the members of the families $\mathcal{B}^{(j_r)}$, $r = 1,\ldots,s$, we shall show that there is a B^* which contains at least

(1.10.5) $$\exp\{c_2\sqrt{n} \log n\}$$

B's, where $c_2 = \frac{1}{2} c_6$.

Suppose that for sufficiently large n, (1.10.5) is false. Let $\mathcal{I}^{(j_r)}$ be the family of subsets of S containing at least $\exp\{c_2\sqrt{n} \log n\}$ sets B. Then $\mathcal{I}^{(j_r)} \cap \mathcal{B}^{(j_r)} = \phi$. Set $\mathcal{V}^{(j_r)} = \mathcal{I}^{(j_r)} \cup \mathcal{B}^{(j_r)}$. Then

(1.10.6) $$|\mathcal{V}^{(j_r)}| = |\mathcal{I}^{(j_r)}| + |\mathcal{B}^{(j_r)}| > |\mathcal{I}^{(j_r)}| + \frac{1}{2} c_1\binom{n}{j_r} .$$

The theorem will be proven by showing

(1.10.7) $$|\mathcal{V}^{(j_s)}| > \binom{n}{j_s} .$$

To obtain this contradiction, we will find a suitable lower bound for $|V^{(j_r)}|$ for all $r \leq s$. Specifically we show that for $r \leq s$

$$(1.10.8) \qquad |V^{(j_r)}| > (r + o(1))\,\frac{1}{2}\,c_1 \binom{n}{j_r}$$

For $r = 1$, we clearly have

$$|V^{(j_1)}| \geq |B^{(j_1)}| > \frac{1}{2}\,c_1 \binom{n}{j_1} \quad .$$

Suppose (1.10.8) is true for $r - 1$.

Now there are $|V^{(j_{r-1})}| \binom{n - j_{r-1}}{j_r - j_{r-1}}$ ways in which some subset of S having j_r elements can contain some member of $V^{(j_{r-1})}$ since there are $\binom{n - j_{r-1}}{j_r - j_{r-1}}$ subsets of S with j_r elements containing a given member of $V^{(j_{r-1})}$. The members of $I^{(j_r)}$ each contain at most $\binom{j_r}{j_{r-1}}$ members of $V^{(j_{r-1})}$. As for the remaining $\binom{n}{j_r} - |I^{(j_{r-1})}|$ subsets of S with j_r elements, since none of them are members of $I^{(j_r)}$, none can contain a member of $I^{(j_{r-1})}$. Also, since none are members of $I^{(j_r)}$, each contains fewer than $\exp\{c_2\sqrt{n}\,\log n\}$ B's and therefore fewer than $\exp\{c_2\sqrt{n}\,\log n\}$ members of $V^{(j_{r-1})}$. Thus

$$(1.10.9) \quad |V^{(j_{r-1})}| \binom{n - j_{r-1}}{j_r - j_{r-1}} < |I^{(j_r)}| \binom{j_r}{j_{r-1}} + \binom{n}{j_r} \exp\{c_2\sqrt{n}\,\log n\}$$

or

$$|I^{(j_r)}| > |V^{(j_{r-1})}| \binom{n - j_{r-1}}{j_r - j_{r-1}} \binom{j_r}{j_{r-1}}^{-1}$$

$$- \binom{n}{j_r} \exp\{c_2\sqrt{n} \log n\} \binom{j_r}{j_{r-1}}^{-1} .$$

Using (1.10.4) with $c_2 = \frac{1}{2} c_6$ and

$$\binom{n - j_{r-1}}{j_r - j_{r-1}} \binom{j_r}{j_{r-1}}^{-1} = \binom{n}{j_{r-1}}^{-1} \binom{n}{j_r}$$

we obtain

$$|I^{(j_r)}| > |V^{(j_{r-1})}| \binom{n}{j_{r-1}}^{-1} \binom{n}{j_r} - \binom{n}{j_r} \exp\{-c_2\sqrt{n} \log n\} . \tag{1.10.10}$$

Since (1.10.8) is assumed to hold for $r - 1$,

$$|I^{(j_r)}| > (r - 1 + o(1)) \frac{1}{2} c_1 \binom{n}{j_r} . \tag{1.10.11}$$

(1.10.11) and (1.10.6) give (1.10.8) for r, and therefore for all $r \leq s$. But for $r = s$, (1.10.8) implies (1.10.7), and this contradiction proves Theorem 1.10.

CHAPTER II

PRIMITIVE SEQUENCES

The study of primitive sequences has been concerned chiefly with the determination of the density of such sequences and the behaviour of the sum $\sum_{a_i \le n} \frac{1}{a_i}$, $\{a_i\}$ a primitive sequence, as n tends to infinity. In this chapter, we shall obtain results in these areas.

§ 2.1 Asymptotic and logarithmic density of primitive sequences.

Given an infinite primitive sequence $A = \{a_i\}$, it can easily be seen, by considering the associated sequence $\{m_i\}$, where m_i is the largest odd divisor of a_i , that the density of A is not greater than 1/2. Behrend [2], however, proved that no infinite primitive sequence can have $\overline{d}A = 1/2$.

<u>Theorem 2.1</u> If A is an infinite primitive sequence, then $\overline{d}A < 1/2$.

<u>Proof</u>. Let 2^α be the largest power of 2 dividing all $a_i \in A$ and let k be the least integer for which $2^\alpha \| a_k$.

For a given x and each a_i let $\alpha_i = [(\log x - \log a_i)/\log 2]$ and let M_x be defined by

$$M_x = \{a_i 2^{\alpha_i} : i \neq k,\ a_i \le x\} \cup \{a_k u : u \text{ odd}, \frac{x}{2a_k} \le u \le \frac{x}{a_k}\} .$$

The elements of M_x are all distinct, for if $a_i 2^{\alpha_i} = a_j 2^{\alpha_j}$, (we can assume $\alpha_j \geq \alpha_i$) then $a_j | a_i$. If $a_i 2^{\alpha_i} = a_k u$, then since $2^{\alpha+1} \nmid a_k u$ but $2^{\alpha} | a_i$, we must have $\alpha_i = 0$ and $a_k | a_i$. Also, if $m \in M_x$ then $\frac{x}{2} \leq m \leq x$. This is clear if $m = a_k u$, since $\frac{x}{2a_k} \leq u \leq \frac{x}{a_k}$. If $m = a_i 2^{\alpha_i}$, the inequality

$$\frac{\log x - \log a_i}{\log 2} - 1 < \alpha_i \leq \frac{\log x - \log a_i}{\log 2}$$

gives $\frac{x}{2} \leq a_i 2^{\alpha_i} \leq x$.

Now M_x contains $A(x) + \frac{x}{4a_k} + O(1)$ elements, so that

$$A(x) + \frac{x}{4a_k} \leq \frac{x}{2} + O(1) .$$

Thus
$$\bar{d}A = \limsup_{x \to \infty} \frac{A(x)}{x} \leq \frac{1}{2} - \frac{1}{4a_k} < \frac{1}{2} .$$

While any primitive sequence must have upper density less than 1/2, there are sequences whose upper density is arbitrarily close to 1/2. We shall construct such a sequence in the following way. Let $\{T_k\}_{k=1}^{\infty}$ be a sequence of integers such that $T_k > 2\, T_{k-1}$ for $k = 2,3,\ldots$. Let $I_{T_k} = \{t : T_k < t \leq 2\, T_k\}$. Although each I_{T_k} is clearly primitive, their union may not be, but we can adjust $\bigcup I_{T_k}$ to obtain such a sequence as follows. Let $I'_{T_1} = I_{T_1}$ and for $k > 1$, let I'_{T_k} be the set obtained from I_{T_k} by deleting all the multiples of the integers in I'_{T_j} for $j < k$. Then $\bigcup I'_{T_k}$ is primitive. We shall show that if T_1 is large enough and if the T_k

increase rapidly enough, then the upper density of $\bigcup I'_{T_k}$ is close to 1/2.

We first prove a theorem of Erdös [6]. In what follows, we denote by N the set of natural numbers.

Theorem 2.2 Let $I_T = \{t : T < t \leq 2\,T\}$ and let $B(I_T) = \{nt : n \in N,\ t \in I_T\}$. Then $\lim_{T\to\infty} d\,B(I_T) = 0$.

Proof. Let $\varepsilon > 0$ be given. Denote by $N^{(1)}$ the set $\{n \in N : \Omega_T(n) \leq \frac{2}{3} \text{loglog } T\}$ and by $N^{(2)}$ the set $\{n \in N : \Omega_T(n) > \frac{2}{3} \text{loglog } T\}$, where $\Omega_T(n)$ is the number of prime power divisors p^α of n with $p \leq T$.

We will prove the theorem by applying Corollary 1.5 , namely that for our given ε, and $n > T > T_0(\varepsilon)$, the number of $u \leq n$ for which $|\Omega_T(u) - \text{loglog } T| \geq \frac{1}{3} \text{loglog } T$ is less than εT .

Let $E(n) = B(I_T) \cap [1, n]$. Each number e in E(n) is of one of three types:

Type 1 $e = t^{(1)}m$ where $t^{(1)} \in N^{(1)} \cap I_T$. If T is large enough, there are at most $2\,T\varepsilon$ such $t^{(1)}$'s . Therefore E(n) has at most

$$\sum_{t^{(1)} \in N^{(1)} \cap I_T} \frac{n}{t^{(1)}} < \frac{n}{T} \sum_{t^{(1)} \in N^{(1)} \cap I_T} 1 < 2\,\varepsilon\, n$$

numbers of this type.

Type 2 $e = t^{(2)} m^{(1)}$ where $t^{(2)} \in N^{(2)} \cap I_T$ and $m^{(1)} \in N^{(1)}$. There are at most

$$\sum_{t \in I_T} \sum_{m^{(1)} < \frac{n}{t}} 1 < \sum_{t \in I_T} \frac{n\varepsilon}{t} < n\varepsilon$$

numbers of this type in $E(n)$ if n is sufficiently large, say $n > 2T^2$.

Type 3 $e = t^{(2)} m^{(2)}$ where $t^{(2)} \in N^{(2)} \cap I_T$ and $m^{(2)} \in N^{(2)}$. Since $\Omega_T(t^{(2)} m^{(2)}) > \frac{4}{3} \log\log T$ there are at most εn numbers of this type in $E(n)$.

Hence $|E(n)| < 2\varepsilon n + \varepsilon n + \varepsilon n = 4\varepsilon n$ for $T > T_0(\varepsilon)$ and $n > 2T^2$. Since ε is arbitrary, the result follows.

Theorem 2.3 (Besicovitch [3]) For any $\varepsilon > 0$ there is primitive sequence A with $\overline{d}A > \frac{1}{2} - \varepsilon$.

Proof. The set $B(I_T)$ is the union of congruence classes mod $(2T)!$. Thus for $m > (2T)!$ any set of m consecutive integers contains at most $d\,B(I_T)\,2\,m$ members of $B(I_T)$.

Let $\varepsilon > 0$ and set $\varepsilon_k = \left(\frac{1}{2}\right)^k \varepsilon$. By Theorem 2.2 a sequence $\{T_1, T_2, \ldots\}$ of integers can be chosen so that $d\,B(I_{T_k}) < \varepsilon_k$ and $T_{k+1} > (2T_k)!$. As we saw previously, a primitive sequence A may be constructed from $\bigcup_{k=1}^{\infty} I_{T_k}$ by taking $A = \bigcup_{k=1}^{\infty} I'_{T_k}$ where $I'_{T_k} = I_{T_k} \sim \bigcup_{r=1}^{k-1} B(I_{T_r})$. For each k, I_{T_k} has at least

$$T_k - \sum_{r=1}^{k-1} dB(I_{T_r})\, 2T_k > T_k(1 - 2\varepsilon)$$

elements so that

$$\overline{d}A \geq \lim_{k\to\infty} T_k(1-2\varepsilon)/(2T_k) = \frac{1}{2} - \varepsilon .$$

The primitive sequence A constructed in the preceding theorem has the property that, while $A(n) > (\frac{1}{2} - \varepsilon)\, n$ for infinitely many values of n, $A(n)/n$ is also arbitrarily small infinitely often. This suggests that the lower density of any primitive sequence A may be 0. More than this is true, however; we will prove that $\overline{\delta}A = 0$.

<u>Theorem 2.4</u> If A is a primitive sequence, then $\sum_{a_i \leq n} \frac{1}{a_i \log a_i} = O(1)$. Equivalently, $\sum_{a_i \varepsilon A} \frac{1}{a_i \log a_i}$ converges.

<u>Proof</u>. For $a_i \varepsilon A$, let p_i be the largest prime factor of a_i. Let $S^{(i)}$ be the sequence of natural numbers all of whose prime factors are larger than p_i. Let $R^{(i)} = a_i S^{(i)}$, that is, $R^{(i)}$ is obtained from $S^{(i)}$ by multiplying each number in $S^{(i)}$ by a_i.

For $i \neq j$, $R^{(i)} \cap R^{(j)} = \phi$ since if $a_i s_1^{(i)} = a_j s_2^{(j)}$, (we can assume $p_i < p_j$) then $a_i | a_j$. Hence, for all n, $\sum_{i=1}^{n} dR^{(i)} \leq 1$.

Now $dR^{(i)} = \frac{1}{a_i} dS^{(i)} = \frac{1}{a_i} \prod_{p \leq p_i} \left(1 - \frac{1}{p}\right)$ so that, on letting $n \to \infty$, we get $\sum_{i=1}^{\infty} \frac{1}{a_i} \prod_{p \leq p_i} \left(1 - \frac{1}{p}\right) \leq 1$. Since $\prod_{p \leq p_i} \left(1 - \frac{1}{p}\right) > \frac{c}{\log p_i} \geq \frac{c}{\log a_i}$, by Theorem 1.2, we conclude that

$$\sum_{i=1}^{\infty} \frac{1}{a_i \log a_i} = O(1) .$$

<u>Theorem 2.5</u> Every primitive sequence has logarithmic density 0.

<u>Proof</u>. It is clearly enough to show that $\overline{\delta}A = 0$.

Let $n > N \geq 1$. Then

$$\frac{1}{\log n} \sum_{a_i \leq n} \frac{1}{a_i} = \frac{1}{\log n} \sum_{a_i \leq N} \frac{1}{a_i} + \frac{1}{\log n} \sum_{N < a_i \leq n} \frac{1}{a_i}$$

so that $$\overline{\delta}A = \limsup_{n \to \infty} \frac{1}{\log n} \sum_{a_i \leq n} \frac{1}{a_i} = \limsup_{n \to \infty} \frac{1}{\log n} \sum_{N < a_i \leq n} \frac{1}{a_i} .$$

Now $$\frac{1}{\log n} \sum_{N < a_i \leq n} \frac{1}{a_i} \leq \sum_{N < a_i \leq n} \frac{1}{a_i \log a_i} < \sum_{N < a_i} \frac{1}{a_i \log a_i}$$ and by

Theorem 2.4 $$\lim_{N \to \infty} \sum_{N < a_i} \frac{1}{a_i \log a_i} = 0 .$$

Thus $\overline{\delta}A = 0$.

§ 2.2 The Behrend-Pillai Theorems.

By Theorem 2.5, all primitive sequences have logarithmic density 0. An equivalent statement to this is that

$$\frac{1}{\log n} \sum_{a_i \leq n} \frac{1}{a_i} = o(1) \quad \text{as} \quad n \to \infty$$

for every primitive sequence $A = \{a_i\}$. We discuss in this section the Behrend-Pillai theorems which give much more accurate information about the behaviour of $\sum_{a_i \leq n} \frac{1}{a_i}$.

<u>Theorem 2.6</u> (Behrend [2]) There exists a constant c such that for any primitive sequence $A = \{a_i\}$

$$\frac{1}{\log n} \sum_{a_i \leq n} \frac{1}{a_i} \leq c \frac{1}{\sqrt{\log\log n}} \quad , \quad n \geq 3 .$$

<u>Proof</u>. For any natural number u, let $d_1(u)$ be the number of divisors of u which are also members of A. Then

$$\sum_{u \leq n} d_1(u) = \sum_{ma_i \leq n} 1 = \sum_{a_i \leq n} \left[\frac{n}{a_i}\right]$$

$$= n \sum_{a_i \leq n} \frac{1}{a_i} + O(n) .$$

Hence

(2.6.1) $$\sum_{a_i \leq n} \frac{1}{a_i} = \frac{1}{n} \sum_{u \leq n} d_1(u) + O(1) .$$

Behrend obtained his result by finding a bound for $\sum_{u \leq n} d_1(u)$.

Let $d_2(u)$ be the number of elements in a maximal primitive set of divisors of u; that is, a maximal set of divisors, no one of which divides any other. Then $d_1(u) \leq d_2(u)$. We shall first consider primitive sequences all of whose terms are square-free. In this case $d_1(u) \leq d_2(v)$ where v is the square-free part of u.

So let us suppose v is square-free. As detailed in Chapter I, we can make a correspondence between the prime factors of v and the elements of any set S of order $\omega(v)$. Under this correspondence, any primitive set of divisors of u will be associated with a primitive family of subsets of the set S.

By Sperner's Theorem, every primitive family of subsets of S has at most $\binom{\omega(v)}{[\omega(v)/2]}$ elements. Thus

$$d_2(v) \leq \binom{\omega(v)}{[\omega(v)/2]}$$

and for any u

$$d_1(u) \leq \binom{\omega(u)}{[\omega(u)/2]} \ .$$

By Stirling's formula,

$$\binom{n}{[n/2]} \leq c_1 \frac{2^n}{\sqrt{n}}$$

Therefore, by (2.6.1),

$$\sum_{a_i \leq n} \frac{1}{a_i} \leq c_1 \frac{1}{n} \sum_{u \leq n} \frac{2^{\omega(u)}}{\sqrt{\omega(u)}} + 0(1) \ .$$

For any $k > 1$,

$$\sum_{u \leq n} \frac{2^{\omega(u)}}{\sqrt{\omega(u)}} = \sum_{\substack{u \leq n \\ \omega(u) \leq k}} \frac{2^{\omega(u)}}{\sqrt{\omega(u)}} + \sum_{\substack{u \leq n \\ \omega(u) > k}} \frac{2^{\omega(u)}}{\sqrt{\omega(u)}}$$

$$\leq n \frac{2^k}{\sqrt{k}} + \frac{1}{\sqrt{k}} \sum_{u \leq n} 2^{\omega(u)} \ ,$$

Now $2^{\omega(u)} \leq d(u)$ and $\sum_{u \leq n} d(u) \leq 2n \log n$, where $d(u)$ is the

number of divisors of u.

Thus
$$\sum_{u \le n} \frac{2^{\omega(u)}}{\sqrt{\omega(u)}} \le n \frac{2^k}{\sqrt{k}} + \frac{2}{\sqrt{k}} n \log n \ .$$

For $n \le 3$, choosing $k = [\log\log n]$ gives

$$2^k < \log n \qquad \text{and} \qquad \sum_{u \le n} \frac{2^{\omega(u)}}{\sqrt{\omega(u)}} \le 3 \frac{n \log n}{\sqrt{\log\log n}} \ .$$

Hence
$$\sum_{a_i \le n} \frac{1}{a_i} \le c_2 \frac{\log n}{\sqrt{\log\log n}} \qquad \text{for } n \ge 3 \ ,$$

for primitive sequences A whose terms are square-free.

To prove the theorem in the general case, let A be any primitive sequence, and $\{a_i^{(k)}\}$ the subsequence of A all of whose terms have greatest square factor k^2. For each $i \in N$, $a_i^{(k)} = k^2 q_i^{(k)}$ where $q_i^{(k)}$ is square free. Thus

$$\sum_{a_i \le n} \frac{1}{a_i} = \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{q_i^{(k)} \le \frac{n}{k^2}} \frac{1}{q_i^{(k)}}$$

$$\le \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{q_i^{(k)} \le n} \frac{1}{q_i^{(k)}} \ .$$

By what has been proven,

$$\sum_{q_i^{(k)} \le n} \frac{1}{q_i^{(k)}} \le c_2 \frac{\log n}{\sqrt{\log\log n}} \ .$$

Therefore $$\sum_{a_i \le n} \frac{1}{a_i} \le c_2 \frac{\log n}{\sqrt{\log\log n}} \sum_{k=1}^{\infty} \frac{1}{k^2}$$

$$= c \frac{\log n}{\sqrt{\log\log n}} .$$

It is also possible to bound sup $\sum_{a_i \le n} \frac{1}{a_i}$ from below, where the supremum is taken over all finite primitive sequences $A \subsetneq \{1,2,\ldots,n\}$. This was first done by Pillai [12] .

<u>Theorem 2.7</u> There is a $c' > 0$, such that, for any $n \ge 3$, there exists a primitive sequence $A' \subsetneq \{1,2,\ldots,n\}$ for which

(2.7.1) $$\frac{1}{\log n} \sum_{a \varepsilon A} \frac{1}{a} \ge c' \frac{1}{\sqrt{\log\log n}} .$$

<u>Proof</u>. For $r \ge 1$ let $\{b_i^{(r)}\}$ denote the sequence of square-free integers having exactly r distinct prime factors. This sequence is clearly primitive. We shall prove that if $r = [\log\log n]$ then the sequence A' of all the terms of $b_i^{(r)}$ which are not greater than n satisfies (2.7.1).

By (1.1.1) and (1.1.2) we can choose c_1 so that, for $n > c_1$,

(2.7.2) $$\sum_{p \le n} \frac{1}{p} \ge \log\log n$$

and

(2.7.3) $$\sum_{p \le n} \frac{\log p}{p} \le 2 \log n .$$

Since it is sufficient to prove (2.7.1) for large n, we shall assume

(2.7.4) $$\log\log n > 8 .$$

Also, we shall assume that

(2.7.5) $$r \leq 2(\log\log n - 4) .$$

Lemma 2.8 Let $\lambda_r(n) = \sum_{b_i^{(r)} \leq n} \frac{1}{b_i^{(r)}}$.

Then $$\lambda_r(n) \geq \frac{(\log\log n - 4)^r}{r!}$$

provided that $n > c_1^{2^{r-1}}$.

Proof. By (2.7.2) the lemma is true for $r = 1$. Let us suppose it is true for $r \geq 1$ and let $n > c_1^{2^r} = c_1^{2^{(r+1)-1}}$.

Then $$(r+1)\, \lambda_{r+1}(n) \geq \sum_{pb_i^{(r)} \leq n} \frac{1}{pb_i^{(r)}} \geq \sum_{p \leq \sqrt{n}} \frac{1}{p} \lambda_r\left(\frac{n}{p}\right) .$$

Now $p \leq \sqrt{n}$ implies $\frac{n}{p} > c_1^{2^{r-1}}$, so that, by the induction hypothesis,

$$(r+1)\, \lambda_{r+1}(n) \geq \frac{1}{r!} \sum_{p < \sqrt{n}} \frac{1}{p} \left(\log\log \frac{n}{p} - 4\right)^r$$

$$= \frac{(\log\log n - 4)^r}{r!} \tau_r(n)$$

where $$\tau_r(n) = \sum_{p \leq \sqrt{n}} \frac{1}{p} \left(1 + \frac{\log(1 - (\log p)/\log n)}{\log\log n - 4}\right)^r .$$

We complete the proof by induction by showing that

$$\tau_r(n) \geq \text{loglog} - 4 \qquad \text{for} \quad n > c_1^{2^n} .$$

We have $p \leq \sqrt{n}$ so that

$$0 < -\log\left(1 - \frac{\log p}{\log n}\right) \leq (2 \log 2) \frac{\log p}{\log n} .$$

Therefore $$\left(1 + \frac{\log(1 - (\log p)/\log n)}{\text{loglog } n - 4}\right)^r \geq \left(1 - \frac{(2 \log 2)(\log p)/\log n}{\text{loglog } n - 4}\right)^r$$

$$\geq 1 - \frac{(2 \log 2)\, r\, (\log p)/\log n}{\text{loglog } n - 4}$$

$$\geq 1 - (4 \log 2)\ (\log p)/\log n$$

by (2.7.5).

Thus $$\tau_r(n) \geq \sum_{p \leq \sqrt{n}} \frac{1}{p} - \frac{4 \log 2}{\log n} \sum_{p \leq \sqrt{n}} \frac{\log p}{p} ,$$

and, as $x > c_1^{2^r} > c_1^2$,

$$\tau_r(n) \geq \text{loglog } n - \log 2 - 4 \log 2$$

$$> \text{loglog } n - 4$$

by (2.7.2) and (2.7.3), proving the lemma.

To complete the proof of the theorem, we let n be so large that $n > c_1^{2^{\text{loglog } n - 1}}$. Taking $r = [\text{loglog } n]$, we have $n > c_1^{2^{r-1}}$ and (2.7.5) holds because of (2.7.4). Therefore by Lemma 2.8,

$$\lambda_r(n) \geq \frac{(\text{loglog } n - 4)^{[\text{loglog } n]}}{[\text{loglog } n]!}$$

Using Stirling's formula to estimate the factorial, we obtain

$$\lambda_r(n) = \sum_{b_i^{(r)} \leq n} \frac{1}{b_i^{(r)}} \geq c' \frac{\log n}{\sqrt{\text{loglog } n}}$$

for some constant $c' > 0$.

By the theorems of Behrend and Pillai,

$\sup \left\{ \sum_{a_i \leq n} \frac{1}{a_i} : \{a_i\} \text{ primitive} \right\}$ is of order $\frac{\log n}{\sqrt{\text{loglog } n}}$. Pillai conjectured the existence of $\limsup_{n \to \infty} \left\{ \frac{\log n}{\sqrt{\text{loglog } n}} \sum_{a \in A} \frac{1}{a} \right\}$

where for each n the supremum is taken over all primitive sequences $A \subseteq \{1,2,\ldots,n\}$. Erdös [5] proved the following result.

<u>Theorem 2.9</u>. Let $\{a_i^{(r)}\}$ be the sequence of integers of degree r, that is, $\Omega(a_i^{(r)}) = r$. Then if $r = [\text{loglog } n]$,

$$\sum_{a_i^{(r)} \leq n} \frac{1}{a_i^{(r)}} = (1 + o(1)) \frac{\log n}{\sqrt{2\pi \text{ loglog } n}} .$$

On the basis of this theorem, he stated that Behrend's method could be modified to show that, for any primitive sequence A,

$$\sum_{a_i \leq n} \frac{1}{a_i} \leq (1 + o(1)) \frac{\log n}{\sqrt{2\pi \text{ loglog } n}}$$

thus claiming to have proved Pillai's conjecture.

However, Anderson [1] replaced Behrend's bound on $d_2(m)$, $d_2(m) \le \binom{\omega(m)}{[\omega(m)/2]}$, for square-free m, by the bound

$$d_2(m) \le \frac{d(m)}{2^{\Omega(m)}} \binom{\Omega(m)}{[\Omega(m)/2]} \qquad \text{for all } m,$$

but even with this new bound he could only show, by Behrend's method, that

$$\sum_{a_i \le n} \frac{1}{a_i} \le (1 + o(1)) \frac{\log n}{\sqrt{\pi} \text{ loglog } n}$$

for any primitive sequence A.

It became clear, therefore, that Erdös' claim had not in fact been justified and that in order to settle Pillai's conjecture, some new ideas would be needed. The real difficulty, in view of Anderson's observations lies in handling the "non square-free" case. This difficulty was finally overcome by Erdös, Sárközi, and Szemerédi [7]. They were able to show that if $\{a_i\}$ is primitive, then in the sum $\sum_{a_i \le n} \frac{1}{a_i}$, the contribution from terms with a "large" square part is "small". We now present their argument in detail.

<u>Theorem 2.10</u> For any primitive sequence A

$$\sum_{a_i \le n} \frac{1}{a_i} \le (1 + o(1)) \frac{\log n}{\sqrt{2\pi} \text{ loglog } n} .$$

Proof. Set $\sum\limits_{x}^{(n)} = \sum\limits_{\substack{t\le n \\ \Omega(t)=x}} \frac{1}{t}$.

By Theorem 2.9, $\sum\limits_{r}^{(n)} = (1 + o(1)) \dfrac{\log n}{\sqrt{2\pi \operatorname{loglog} n}}$, where $r = [\operatorname{loglog} n]$.

To prove the theorem, we shall show that, for any primitive sequence A,

(2.10.1) $$\sum_{a_i \le n} \frac{1}{a_i} \le (1 + o(1)) \sum_{r}^{(n)} .$$

We first prove a lemma which shows that in the sum $\sum\limits_{a_i\le n} \frac{1}{a_i}$, all terms with a "large" square part may be ignored. Or more precisely:

<u>Lemma 2.11</u> $\sum_o \frac{1}{t} < (1 + o(1)) \dfrac{\log n}{r^2 \log(3/2)}$ where in $\sum_o$, $1 \le t \le n$ and $\Omega(t) - \omega(t) > 8 \log r$, with $r = [\operatorname{loglog} n]$.

<u>Proof</u>. Suppose $\Omega(t) - \omega(t) > 8 \log r$. Put $t = a^2T$ where T is square-free. We claim $\Omega(a^2) - \omega(a^2) > 4 \log r$. In order to see this suppose $t = p_1^{\alpha_1}\cdots p_k^{\alpha_k}$ where $\alpha_1,\ldots,\alpha_s \ge 2$ while $\alpha_{s+1} = \cdots = \alpha_k = 1$. Then

$$\Omega(t) - \omega(t) = (\alpha_1 - 1) + \cdots + (\alpha_s - 1) > 8 \log r .$$

For $i = 1,\ldots,s$ put $\alpha_i = 2\beta_i + \varepsilon_i$ where $\varepsilon_i = 0$ or 1. We consider two cases.

<u>Case 1</u>. $s > 4 \log r$

Then $$\Omega(a^2) - \omega(a^2) = \sum_{i=1}^{s} 2\beta_i - 1 \ge s > 4 \log r .$$

Case 2. $s \le 4 \log r$

$$\Omega(a^2) - \omega(a^2) = (2\beta_1 - 1) + \dots + (2\beta_s - 1)$$

$$= (\alpha_1 - 1) + \dots + (\alpha_s - 1) - (\varepsilon_1 + \dots + \varepsilon_s)$$

$$> 8 \log r - 4 \log r = 4 \log r .$$

Now $$\sum\nolimits_o \frac{1}{t} \le \sum_{\substack{1 \le a^2T \le n \\ \Omega(a^2)-\omega(a^2)>4\log r}} \frac{1}{a^2T}$$

$$\le \sum_{\substack{1 \le a^2 \le n \\ \Omega(a^2)-\omega(a^2)>4\log r}} \frac{1}{a^2} \sum_{\substack{1 \le T \le n \\ T \text{ square-free}}} \frac{1}{T}$$

$$< \sum_p \left(\frac{1}{p^2}\right)^{2\log r} \sum_{1 \le t \le n} \frac{1}{t} ,$$

since $\Omega(a^2) - \omega(a^2) > 4 \log r$ implies $\Omega(a) > 2 \log r$.

Now $$\sum_p \left(\frac{1}{p^2}\right)^{2\log r} \sum_{1 \le t \le n} \frac{1}{t}$$

$$< \left(\frac{2}{3}\right)^{2\log r} (1 + o(1)) \log n$$

$$= (1 + o(1)) \frac{\log n}{r^{2\log 3/2}}$$

which proves the lemma. For the theorem, it will be enough to know that

$$\sum\nolimits_o \frac{1}{t} = o\left(\frac{\log n}{\sqrt{r}}\right) .$$

Let A be any primitive sequence for which $\Omega(a_i) - \omega(a_i) \le 8 \log r$ for all i. We shall show that (2.10.1) holds for such a sequence, and this, together with the lemma, will establish the theorem.

Now

$$(2.10.2) \quad \sum_{a_i \le n} \frac{1}{a_i} = \sum_{s>r} \sum_{a_j^{(s)} \le n} \frac{1}{a_j^{(s)}} + \sum_{a_j^{(r)} \le n} \frac{1}{a_j^{(r)}} + \sum_{s<r} \sum_{a_j^{(s)} \le n} \frac{1}{a_j^{(s)}},$$

$$= \sum\nolimits_1 + \sum\nolimits_2 + \sum\nolimits_3 ,$$

where $a_j^{(s)}$ means that $\Omega(a_j^{(s)}) = s$.

For $s > r$, we shall replace each $a_j^{(s)}$ by all its divisors b of degree r to obtain a sequence $\{b_i\}$ where $\Omega(b_i) = r$ and for each i, $b_i | a_j^{(s)}$ for some j. Similarly, for $r > s$, we shall obtain a sequence $\{d_i\}$, where $\Omega(d_i) = r$ and for each i, $a_j^{(s)} | d_i$ for some j. Since $a_1 < \ldots < a_k \le n$ is a primitive sequence, the sequences $\{b_i\}$, $\{a_j^{(r)}\}$, and $\{d_i\}$ are disjoint, hence

$$(2.1.3) \quad \sum_{b_i \le n} \frac{1}{b_i} + \sum_{a_i^{(r)} \le n} \frac{1}{a_{i_j}^{(r)}} + \sum_{d_i \le n} \frac{1}{d_i} \le \sum_{r}^{(n)} .$$

(2.10.1) will follow from

$$(2.10.4) \quad \sum\nolimits_1 \le (1 + o(1)) \sum_{b_i \le n} \frac{1}{b_i}$$

and

$$(2.10.5) \quad \sum\nolimits_3 \le (1 + o(1)) \sum_{d_i \le n} \frac{1}{d_i} + o\left(\frac{\log n}{\sqrt{r}}\right)$$

by (2.10.2) and (2.10.3).

We first prove (2.10.4). Set $s_1 = \max_i \{\Omega(a_i)\}$. We can assume $s_1 > r$, otherwise (2.10.4) is trivial. We transform the $a_j^{(s)}$'s $(s > r)$ into the b's by an induction process in such a way that (2.10.4) holds.

Consider the set

$$U_1 = \{u_i^{(s_1-1)} : u_i^{(s_1-1)} \mid a_j^{(s_1)} \text{ for some } a_j^{(s_1)}\} .$$

Since $\{a_i\}$ is a primitive sequence, no $u_i^{(s_1-1)}$ is an $a_j^{(s_1-1)}$. Now consider

$$U_2 = \{u_i^{(s_1-2)} : u_i^{(s_1-2)} \mid a_j^{(s_1-1)} \text{ for some } a_j^{(s_1-1)} \text{ or }$$
$$u_i^{(s_1-2)} \mid u_i^{(s_1-1)} \text{ for some } u_i^{(s_1-1)}\} .$$

Again, no $u_i^{(s_1-2)}$ is an $a_j^{(s_1-2)}$. Repeating this construction $s_1 - r$ times gives the set U_{s_1-r} which is the set of b's.

We have

$$\sum_{u_i^{(s_1-1)} \le n} \frac{1}{u_i^{(s_1-1)}} \sum_{p \le n} \frac{1}{p} \ge \omega(a_j^{(s_1)}) \sum_{a_j^{(s_1)} \le n} \frac{1}{a_j^{(s_1)}}$$

and, for $1 < m \le s_1 - r$,

$$\sum_{u_i^{(s_1-m)} \le n} \frac{1}{u_i^{(s_1-m)}} \sum_{p \le n} \frac{1}{p} \ge \omega(a_j^{(s_1-m+1)}) \sum_{a_j^{(s_1-m+1)} \le n} \frac{1}{a_j^{(s_1-m+1)}}$$

$$+ \omega(u_i^{(s_1-m+1)}) \sum_{u_i^{(s_1-m+1)} \le n} \frac{1}{u_i^{(s_1-m+1)}} .$$

By Lemma 2.11,

$$\omega(u_i^{(s_1-m+1)}) \ge s_1 - m + 1 - 8 \log r ,$$

and since $u_i^{(s_1-m+1)}$ divides some $a_j^{(s)}$

$$\omega(a_j^{(s_1-m+1)}) \ge s_1 - m + 1 - 8 \log r .$$

Now $\displaystyle\sum_{p \le n} \frac{1}{p} < \text{loglog } n + C = r + c$ so that

$$(2.10.6) \quad \sum \frac{1}{u_i^{(s_1-m)}} \ge \frac{s_1-m+1-8\log r}{r+c} \left(\sum \frac{1}{u_i^{(s_1-m+1)}} + \sum \frac{1}{a_j^{(s_1-m+1)}} \right) .$$

We have

$$(2.10.7) \quad \frac{s_1 - m + 1 - 8 \log r}{r + c} > 1 \quad \text{for} \quad s_1 - m > r + 9 \log r$$

and

$$(2.10.8) \quad \frac{s_1 - m + 1 - 8 \log r}{r + c} > 1 - \frac{9 \log r}{r} \quad \text{for} \quad s_1 - m > r \ge r_0 .$$

Thus (2.10.6), (2.10.7) and (2.10.8) give

$$\sum_{b_i \le n} \frac{1}{b_i} = \sum_{u_i^{(r)} \le n} \frac{1}{u_i^{(r)}} > \left(1 - \frac{9 \log r}{r}\right)^{9 \log r} \sum_{s > r} \sum_{a_j^{(s)} \le n} \frac{1}{a_j^{(s)}}$$

$$= (1 + o(1)) \sum\nolimits_1 ,$$

which proves (2.10.4).

We now prove (2.10.5). Setting $s_2 = \min_i\{\Omega(a_i)\}$, we can assume $s_2 < r$. Starting with the integers $a_j^{(s_2)}$, we replace them by the integers $u_i^{(s_2+1)}$ of the form $pa_j^{(s_2)}$ where $p < n^{\frac{1}{r^2}}$. Repeating this procedure $r - s_2$ times gives the integers $u_i^{(r)}$ which consist of some or all of the d_i's and some or all of the integers in the interval $[n, n^{1+\frac{1}{r}}]$ having r prime factors, (since $u_i^{(r)} < n\, n^{\frac{1}{r^2}(r-s_2)} \le n^{1+\frac{1}{r}}$.) .

Each integer having $s_2 + m + 1$ prime factors has at most $s_2 + m + 1$ divisors having $s_2 + m$ prime factors, so that

(2.10.10)
$$\sum_{a_j^{(s_2+m)} \le n} \frac{1}{a_j^{(s_2+m)}} + \sum_{u_i^{(s_2+m)} \le n} \frac{1}{u_i^{(s_2+m)}} \sum_{p<n^{\frac{1}{r^2}}} \frac{1}{p}$$

$$\le (s_2 + m + 1) \sum_{u_i^{(s_2+m+1)} \le n} \frac{1}{u_i^{(s_2+m+1)}} .$$

Now $\displaystyle\sum_{p \le n^{\frac{1}{r^2}}} \frac{1}{p} > r - 3 \log r$, so that

$$\sum_{u_i^{(s_1+m+1)} \le n} \frac{1}{u_i^{(s_1+m+1)}} \ge \frac{r - 3 \log r}{s_1 + m + 1} \left(\sum_{u_i^{(s_1+m)} \le n} \frac{1}{u_i^{(s_1+m)}} + \sum_{a_j^{(s_1+m)} \le n} \frac{1}{a_j^{(s_1+m)}} \right) .$$

We have $\dfrac{r - 3 \log r}{s_2 + m + 1} \ge 1$ for $s_2 + m + 1 \le r - 3\log r$

and $\dfrac{r - 3 \log r}{s_1 + m + 1} \ge 1 - \dfrac{3 \log r}{r}$ since $s_2 + m + 1 \le r$.

Hence, by induction on m,

$$(2.10.11)\qquad \sum_{u_i^{(r)}\le n} \frac{1}{u_i^{(r)}} \ge \left(1 - \frac{3\log r}{r}\right)^{3\log r} \sum\nolimits_3 .$$

$$= (1 + o(1)) \sum\nolimits_3 .$$

But

$$(2.10.12)\qquad \sum_{u_i^{(r)}\le n} \frac{1}{u_i^{(r)}} \le \sum_{d_i\le n} \frac{1}{d_i} + \sum_{t=n}^{\left[n^{1+\frac{1}{r}}\right]} \frac{1}{t} = \sum_{d_i\le n} \frac{1}{d_i} + 0\left(\frac{\log n}{r}\right) .$$

Then (2.10.5) follows from (2.10.11) and (2.10.12) and the theorem is proven.

Thus $$\lim_{n\to\infty} \sup_{A} \left\{\left(\frac{\log n}{\sqrt{\log\log n}}\right)^{-1} \sum_{a\in A} \frac{1}{a}\right\} = \frac{1}{\sqrt{2\pi}}$$

where the supremum is taken over all primitive sequences $A \subsetneq \{1,2,\ldots,n\}$.

In view of the result of Pillai one would perhaps expect that there exists an infinite primitive sequence A such that for some constant c

$$\sum_{a_i\le n} \frac{1}{a_i} > c\,\frac{\log n}{\sqrt{\log\log n}} .$$

This, however, has turned out to be not the case. We devote the remainder of this chapter to a discussion of the following result of Erdös, Sárközi, and Szemerédi [8].

Theorem 2.12 Let A be an infinite primitive sequence. Then

$$\sum_{a_i \le x} \frac{1}{a_i} = o\left(\frac{\log x}{\sqrt{\log\log x}}\right) . \tag{2.12.1}$$

Proof. Suppose there is a primitive sequence A for which (2.12.1) does not hold. By the argument used in Behrend's Theorem, we can assume without loss of generality that the terms of A are square-free.

For this sequence, there is a sequence $\{x_i\}$ tending to infinity sufficiently quickly that

$$\sum_{x_{\nu-1} < a_i < x_\nu} \frac{1}{a_i} > c_1 \frac{\log x_\nu}{\sqrt{\log\log x_\nu}} . \tag{2.12.2}$$

That (2.12.2) leads to a contradiction is a consequence of the following theorem.

Theorem 2.13 Let $u < w$ where w is sufficiently large compared to u (say $w > \exp\exp 2u$). Let $a_1 < \ldots < a_k$ be a primitive sequence of square-free integers for which

$$u < a_1 < \ldots < a_k < w \tag{2.13.1}$$

and

$$\sum_{i=1}^{k} \frac{1}{a_i} > c_2 \frac{\log w}{\sqrt{\log\log w}} . \tag{2.13.3}$$

Suppose that y satisfies $y > \exp\exp 2w$ and denote by $b_1 < \dots < b_s \leq y$ the integers of the form $a_i Q_m$, $i = 1,\dots,k$, where all the prime factors of Q_m are greater than u.

Then

$$(2.13.3) \qquad \sum_{i=1}^{s} \frac{1}{b_i} > c_3 \log y \, ,$$

where $c_3 = c_3(c_2)$.

We apply Theorem 2.13 as follows. Let λ be an integer for which $\lambda c_3 > 2$ and let $y > \exp\exp 2x_\lambda$. For each ν, $1 \leq \nu \leq \lambda$, (2.12.2) gives a primitive sequence $A^{(\nu)} = \{a_1^{(\nu)} < \dots < a_{r_\nu}^{(\nu)}\}$, with $x_{\nu-1} < a_1 \dots < a_{r_\nu}^{(\nu)} < x_\nu$, for which (2.13.2) holds with $c_2 = c_1$. Then we have corresponding sequences $B^{(\nu)}$, $\nu = 1,\dots,\lambda$, for which (2.13.3) holds.

Then

$$(2.12.3) \qquad B^{(\nu)} \cap B^{(\nu')} = \phi \quad \text{if} \quad \nu \neq \nu' \, .$$

For if $a_i Q_m^{(\nu)} = b_i^{(\nu)} = b_j^{(\nu')} = a_j Q_n^{(\nu')}$, (we can assume $\nu' > \nu$) then $a_i | a_j Q_n^{(\nu')}$. Since all the prime factors of $Q_n^{(\nu')}$ are greater than $x_{\nu'-1} \geq x_\nu$, $(a_i, Q_n^{(\nu')}) = 1$ and $a_i | a_j$. But $a_i < x_\nu$ and $a_j > x_{\nu'-1} \geq x_\nu$. Therefore $a_i < a_j$, so that $a_i \nmid a_j$.

For $\nu \leq \lambda$, $b_i^{(\nu)} \leq y$ and by (2.13.3),

$$\sum_{i=1}^{s} \frac{1}{b_i^{(\nu)}} > c_3 \log y \, .$$

By (2.12.3) we then have

$$2 \log y > \sum_{t<y} \frac{1}{t} \geq \sum_{\nu=1}^{\lambda} \sum_{i=1}^{s_\nu} \frac{1}{b_i^{(\nu)}} \geq \lambda c_3 \log y > 2 \log y$$

and this contradiction proves Theorem 2.12. Hence we need only prove Theorem 2.13.

<u>Proof of Theorem 2.13</u>. Let $\{t_i\}$ be the sequence of integers all of whose prime factors are greater than w. Then it is easily seen that

$$\sum_{b_i \leq y} \frac{1}{b_i} \geq \sum_{b_i \leq w} \frac{1}{b_i} \sum_{t_i \leq \frac{y}{w}} \frac{1}{t_i} \ .$$

Since $y > \exp \exp 2w$, $\frac{y}{w}$ is large enough that Theorem 1.3 holds, giving

$$\sum_{t_i \leq \frac{y}{w}} \frac{1}{t_i} > c_4 \frac{\log y}{\log w} \ .$$

Thus to prove (2.13.3) we have only to show

(2.13.4) $$\sum_{b_i \leq w} \frac{1}{b_i} > c_5 \log w \ .$$

Let $d_3(n)$ be the number of divisors of n among the a's and $d_4(n)$ the number of divisors of n among the b's. Clearly

$$\sum_{n=1}^{w} d_4(n) = \sum_{i=1}^{s} \left[\frac{w}{b_i}\right] \le w \sum_{i=1}^{s} \frac{1}{b_i} .$$

Thus (2.13.4) will follow from

$$\sum_{n=1}^{w} d_4(n) > c_6 \, w \log w \quad . \tag{2.13.5}$$

Let $n_1 < \ldots < n_t \le w$ be the sequence of integers for which

(2. 3 6) $\omega(n_i) > \log\log w$ and $d_3(n_i) > \frac{c_2}{16} \frac{2^{\omega(n_i)}}{\sqrt{\omega(n_i)}}$.

Now

$$\sum_{i=1}^{t} d_3(n_i) \ge \sum_{n=1}^{w} d_3(n) - \sum\nolimits_1 d_3(n) - \sum\nolimits_2 d_3(n) ,$$

where in $\sum_1$, $n \le w$ and $\omega(n) \le \log\log w$ and in $\sum_2$, $n \le w$,

$\omega(n) > \operatorname{loglog} w$ and $d_3(n) \le \frac{c_2}{16} \frac{2^{\omega(n)}}{\sqrt{\omega(n)}} < \frac{c_2}{16} \frac{2^{\omega(n)}}{\sqrt{\operatorname{loglog} w}}$.

From (2.13.2) we have

$$\sum_{n=1}^{w} d_3(n) \ge w \sum_{i=1}^{k} \frac{1}{a_i} - w > \frac{c_2}{2} \frac{w \log w}{\sqrt{\operatorname{loglog} w}} .$$

Also

$$\sum\nolimits_1 d_3(n) \le w \, 2^{\operatorname{loglog} w} < \frac{c_2}{8} \frac{w \log w}{\sqrt{\operatorname{loglog} w}}$$

and

$$\sum\nolimits_2 d_3(n) < \frac{c_2}{16} \sum_{n=1}^{w} \frac{2^{\omega(n)}}{\sqrt{\log\log w}}$$

$$\le \frac{c_2}{16} \frac{1}{\sqrt{\log\log w}} \sum_{n=1}^{w} d(n)$$

$$< \frac{c_2}{8} \frac{w \log w}{\sqrt{\log\log w}} \ ,$$

since $2^{\omega(n)} \le d(n)$ so that $\sum_{n=1}^{w} 2^{\omega(n)} \le \sum_{n=1}^{w} d(n) < 2\, w \log w$.

Therefore

$$\sum_{i=1}^{t} d_3(n_i) > \frac{c_2}{4} \frac{w \log w}{\sqrt{\log\log w}} \ . \tag{2.13.7}$$

Then (2.13.5) will follow at once from

$$d_4(n_i) > c_7\, d_3(n_i)\sqrt{\omega(n_i)} > c_7\, d_3(n_i)\sqrt{\log\log w} \tag{2.13.8}$$

and (2.13.7), the last inequality of (2.13.8) following from (2.13.6).

To prove (2.13.8) let

$$p_1 < \dots < p_{r_1} \le u < q_1 < \dots < q_{r_2} \le w$$

be the prime factors of n_i. Clearly $r_1 < u$ and by (2.13.6)

$$r_2 > \log\log w - u > u > r_1$$

if $w > \exp\exp 2u$. Also by (2.13.6)

(2.13.9) $$d_3(n_i) > \frac{c_2}{16} \frac{2^{\omega(n_i)}}{\sqrt{\omega(n_i)}} .$$

The product of any a dividing n_i and a number of q's not dividing a is a b dividing n_i, so (2.13.8) will follow from the combinatorial estimate obtained from Theorem 1.9.

To apply Theorem 1.9, we let S be the set of distinct prime factors of n_i, S_1 be the set of p's and S_2 the set of q's. Further, if we identify the a's with the primitive family of subsets A and the b's with the family of subsets B, then since (2.13.9) holds, we have $s = d_4(n_i) > c_9\, 2^{\omega(n_i)}$ which is (2.13.8). Thus Theorem 2.13 and so Theorem 2.12 is proven.

Theorem 2.12 is best possible in the following sense. Suppose that h(x) tends to infinity arbitrarily slowly. Then there is an infinite primitive sequence A for which

$$\limsup_{x \to \infty} \sum_{a_i \le x} \frac{1}{a_i} h(x) \frac{\sqrt{\log\log x}}{\log x} = \infty .$$

To see this, let $\{x_i\}$ be an increasing sequence tending to infinity, and in each interval $(x_{\nu-1}, x_\nu)$ let $A^{(\nu)}$ consist of all the integers with exactly $[\log\log x_\nu]$ distinct prime factors, each of which is larger than $x_{\nu-1}$. By the argument of Pillai, there is a positive constant c such that $\displaystyle\sum_{a_i^{(\nu)} \le x_\nu} \frac{1}{a_i^{(\nu)}} \ge c \frac{\log x_\nu}{\sqrt{\log\log x_\nu}}$

if $x_\nu \to \infty$ sufficiently quickly. Letting $A = \bigcup_{\nu=1}^{\infty} A^{(\nu)}$ gives the desired result.

CHAPTER III

SEQUENCES OF POSITIVE UPPER LOGARITHMIC DENSITY

If a sequence S has positive upper logarithmic density, then by Theorem 2.4, it cannot be primitive. In fact, S must have infinitely many pairs of elements s_i, s_j such that $s_i | s_j$. However this result can be improved substantially in two directions. We shall show that any sequence S of positive upper logarithmic density has a subsequence S', each of whose terms divide the succeeding term, and that, for such a sequence S, $\limsup_{x \to \infty} \frac{f(x)}{x} = \infty$, where $f(x) = \sum_{\substack{s_i < s_j \le x \\ s_i | s_j}} 1$.

Theorem 3.1 (Davenport - Erdös [4]) Let $S = \{s_i\}_{i=1}^{\infty}$ be any sequence with positive upper logarithmic density. Then S has a subsequence $\{s_{i_j}\}_{j=1}^{\infty}$ such that $s_{i_j} | s_{i_{j+1}}$ for all j.

Proof. Suppose $\bar{\delta} S = \alpha > 0$. It is sufficient to show that there is an $s_{i_1} \varepsilon S$ for which

$$(3.1.1) \qquad \limsup_{n \to \infty} \frac{1}{\log n} \sum_{\substack{s_i \le n \\ s_{i_1} | s_i}} \frac{1}{s_i} > 0 ,$$

for then, by repeating the argument with the sequences $S_{(j)} = \{s \varepsilon S : s_{i_j} | s\}$, for each j, we obtain a subsequence $\{s_{i_j}\}_{j=1}^{\infty}$ with the desired property.

If (3.1.1) is false, then for any r and all k, $1 \le k \le r$,

$$\limsup_{n \to \infty} \frac{1}{\log n} \sum_{\substack{s_j \le n \\ s_k | s_j}} \frac{1}{s_j} = 0 .$$

so that

$$\limsup_{n \to \infty} \frac{1}{\log n} \sum_{\substack{s_i \le n \\ s_k \nmid s_i}} \frac{1}{s_i} = \alpha .$$

We shall prove below that for any sequence S, the set of multiples of S, $B(S)$, has logarithmic density $\delta\, B(S)$ and that,

$$\delta\, B(S) = \sum_{i=1}^{\infty} \delta\, B^{(i)}(S) \tag{3.1.2}$$

where $B^{(i)}(S) = \{b \in B(S) : s_i | b \text{ but } s_1 \nmid b, \ldots, s_{i-1} \nmid b\}$.

Since $S \subsetneq B(S)$, we have

$$\alpha \le \limsup_{n \to \infty} \frac{1}{\log n} \sum_{\substack{b_i \le n \\ s_k \nmid b_i}} \frac{1}{b_i}$$

where $1 \le k \le r$ and $b_i \in B(S)$. Hence

$$\alpha \le \delta\, B(S) - \sum_{i=1}^{r} \delta\, B^{(i)}(S)$$

$$= \sum_{i=r+1}^{\infty} \delta\, B^{(i)}(S) .$$

Now by choosing large enough r, we have $\sum_{i=r+1}^{\infty} \delta\, B^{(i)}(S) < \alpha$

by (3.1.2). Thus there must be some s_{i_1} for which (3.1.1) is true.

For any sequence S, we denote by $\mathcal{B}_m(S)$ the set $\mathcal{B}\{s_1,\ldots,s_m\}$; by $\mathcal{B}_m(n)$, the set $\mathcal{B}_m(S)\cap\{1,\ldots,n\}$; and by $\mathcal{B}^{(i)}(n)$ the set $\mathcal{B}^{(i)}(S)\cap\{1,\ldots,n\}$. Then

(3.1.3) $$|\mathcal{B}_m(n)| = \sum_{i=1}^{m} |\mathcal{B}^{(i)}(n)| \ .$$

Now for every i, $|\mathcal{B}^{(i)}(n)| = \left[\frac{n}{s_i}\right] - \sum_{j<i}\left[\frac{n}{[s_i,s_j]}\right] + \sum_{k<j<i}\left[\frac{n}{[s_i,s_j,s_k]}\right] + \ldots$

so that

(3.14) $$d\ \mathcal{B}^{(i)}(S) = \lim_{n\to\infty}\frac{|\mathcal{B}^{(i)}(n)|}{n} = \frac{1}{s_i} - \sum_{j<i}\frac{1}{[s_i,s_j]} + \ldots \ .$$

By (3.1.3) we have

(3.1,5) $$d\ \mathcal{B}_m(S) = \sum_{i=1}^{m} d\ \mathcal{B}^{(i)}(S) \ .$$

Hence $d\ \mathcal{B}^{(i)}(S) \geq 0$ and $d\ \mathcal{B}^{(1)}(S) > 0$ by (3.1.4) and, by (3.1.5),

$$0 < \sum_{i=1}^{m} d\ \mathcal{B}^{(i)}(S) \leq 1 \qquad \text{for all m.}$$

Setting

(3.1.6) $$\beta = \sum_{i=1}^{\infty} d\ \mathcal{B}^{(i)}(S) \ ,$$

we have

(3.1.7) $$0 < \beta \leq 1 \ ,$$

and by (3.1.5)

(3.1.8) $$\beta = \lim_{m \to \infty} d\, \mathcal{B}_m(\mathcal{S}) \ .$$

As $$\mathcal{B}_m(\mathcal{S}) \subsetneq \mathcal{B}(\mathcal{S}) \ ,$$

(3.1.9) $$\beta \leq \underline{d}\, \mathcal{B}(\mathcal{S}) \ .$$

<u>Lemma 3.2</u> For any sequence $\mathcal{S}$, if $\sum_{i=1}^{\infty} \frac{1}{s_i}$ converges, then $d\, \mathcal{B}(\mathcal{S}) = \lim_{m\to\infty} d\, \mathcal{B}_m(\mathcal{S}) = \beta$.

<u>Proof</u>. $\bar{d}\, \mathcal{B}(\mathcal{S}) \leq d\mathcal{B}_m(\mathcal{S}) + \sum_{i=m+1}^{\infty} \frac{1}{s_i}$. Letting m tend to infinity, gives $\bar{d}\, \mathcal{B}(\mathcal{S}) \leq \beta$, which together with (3.1.9) proves the lemma.

We now complete Theorem 3.1 by proving the following theorem of Davenport and Erdös. We give the elementary proof of Halberstam and Roth [10].

<u>Theorem 3.3</u> For any infinite sequence $\mathcal{S}$, $\mathcal{B}(\mathcal{S})$ possesses logarithmic density, and

$$\delta\, \mathcal{B}(\mathcal{S}) = \underline{d}\, \mathcal{B}(\mathcal{S}) \ .$$

<u>Proof</u>. By (3.1.9), the theorem will be proven by showing that

$$\bar{\delta}\, \mathcal{B}(\mathcal{S}) \leq \beta = \lim_{m \to \infty} d\, \mathcal{B}_m(\mathcal{S}) \ .$$

Let $N^{(k)}$ be the set of natural numbers composed of $p_1,\ldots,p_k$ and put

$$S^{(k)} = S \cap N^{(k)} .$$

Now
$$\sum_{n \in N^{(k)}} \frac{1}{n} = \prod_{i=1}^{k} \left(1 - \frac{1}{p_i}\right)^{-1} = O(\log k) .$$

Therefore $\sum_{i=1}^{\infty} \frac{1}{s_i^{(k)}}$ converges, so by Lemma 3.2, $\mathcal{B}(S^{(k)})$ has asymptotic density.

We now show

$$\lim_{k\to\infty} d\, \mathcal{B}(S^{(k)}) = \beta = \lim_{m\to\infty} d\, \mathcal{B}_m(S) . \tag{3.3.2}$$

Clearly $d\,\mathcal{B}(S^{(k)})$ increases with k, and $d\,\mathcal{B}(S^{(k)}) \le 1$. Hence $\lim_{k\to\infty} d\,\mathcal{B}(S^{(k)})$ exists.

For any r, there is a $k = k(r)$ such that $s_1,\ldots,s_r \in S^{(k)}$. Therefore

$$d\,\mathcal{B}(S^{(k)}) \ge d\,\mathcal{B}_r(S)$$

and, as $r \to \infty$, by (3.1.8),

$$\lim_{k\to\infty} d\,\mathcal{B}(S^{(k)}) \ge \beta . \tag{3.3.3}$$

On the other hand, for any $\varepsilon > 0$ and any k, there is an $r_0 = r_0(\varepsilon,k)$ such that

$$\sum_{i=r+1}^{\infty} \frac{1}{s_i^{(k)}} < \varepsilon \quad \text{for} \quad r > r_0 .$$

Then $$d\,\mathcal{B}(S^{(k)}) \leq d\,\mathcal{B}_r(S^{(k)}) + \sum_{i=r+1}^{\infty} \frac{1}{s_i^{(k)}}$$

$$< d\,\mathcal{B}_r(S^{(k)}) + \varepsilon\ .$$

If h is large enough that

$$\{s_1^{(k)}, \ldots, s_r^{(k)}\} \subseteq \{s_1, \ldots, s_h\}\ ,$$

then $$d\,\mathcal{B}(S^{(k)}) < d\,\mathcal{B}_h(S) + \varepsilon$$

$$< \beta + \varepsilon \quad \text{for all } \varepsilon \text{ and } k.$$

Hence $$\lim_{k\to\infty} d\,\mathcal{B}\, S^{(k)}) \leq \beta + \varepsilon \quad \text{for all} \quad \varepsilon > 0\ ,$$ and thus

(3.3.4) $$\lim_{k\to\infty} d\,\mathcal{B}(S^{(k)}) \leq \beta\ .$$

(3.3.3) and (3.3.4) imply (3.3.2).

If we denote by $M^{(k)}$ the set $\mathcal{B}(S^{(k)}) \cap N^{(k)}$, then any $m^{(k)}$ in $M^{(k)}$ has the form $s^{(k)}n^{(k)}$ and

$$\sum_{i=1}^{\infty} \frac{1}{m_i^{(k)}} = \frac{1}{s_1^{(k)}} \sum_{i=1}^{\infty} \frac{1}{n_i^{(k)}} + \left(\frac{1}{s_2^{(k)}} - \frac{1}{[s_1^{(k)}, s_2^{(k)}]}\right) \sum_{i=1}^{\infty} \frac{1}{n_i^{(k)}} + \cdots\ .$$

Setting $\gamma_k = \sum_{i=1}^{\infty} \frac{1}{n_i^{(k)}} = O(\log k)$, we have

(3.3.5) $$\sum_{i=1}^{\infty} \frac{1}{m_i^{(k)}} = \gamma_k\, d\,\mathcal{B}(S^{(k)})\ .$$

For each k, we set

$$(3.3.6) \qquad \beta(n) = \sum_{b_i \le n} \frac{1}{b_i} = \beta_1(n) + \beta_2(n) \ , \quad \text{where}$$

$$(3.3.7) \qquad \beta_1(n) = \sum_{\substack{b_i \le n \\ b_i \in \mathcal{B}(S^{(k)})}} \frac{1}{b_i} \ , \qquad \beta_2(n) = \sum_{\substack{b_i \le n \\ b_i \notin \mathcal{B}(S^{(k)})}} \frac{1}{b_i} \ .$$

Since $\sum_{i=1}^{\infty} \frac{1}{s_i^{(k)}}$ converges, $d\ \mathcal{B}(S^{(k)})$ exists, so that

$$(3.3.8) \qquad \lim_{n\to\infty} \frac{\beta_1(n)}{\log n} = d\ \mathcal{B}(S^{(k)}) \ .$$

We now consider $\beta_2(n)$. Let h be the unique integer defined by, (p_i is the i'th prime),

$$p_h \le n < p_{h+1} \ .$$

Then
$$\beta_2(n) \le \sum_{\substack{i=1 \\ m_i^{(h)} \notin \mathcal{B}(S^{(k)})}}^{\infty} \frac{1}{m_i^{(h)}}$$

since each $b_i \le n$ is an $m_i^{(h)}$. An $m_i^{(h)}$ contributes to the sum if

$$m_i^{(h)} \in M^{(h)} \sim \bigcup_{i=1}^{\infty} \{u_i M^{(k)} : u_i = p_{k+1}^{\alpha_{k+1}} \cdots p_h^{\alpha_h}\} \ .$$

Thus
$$\beta_2(n) \le \sum_{i=1}^{\infty} \frac{1}{m_i^{(h)}} - \left(\sum_{i=1}^{\infty} \frac{1}{u_i}\right) \sum_{i=1}^{\infty} \frac{1}{m_i^{(k)}} \ .$$

But $\sum_{i=1}^{\infty} \frac{1}{u_i} = \frac{\gamma_h}{\gamma_k}$, so by (3.3.5),

$$\beta_2(n) \le \gamma_h \{d\, B(S^{(h)}) - d\, B(S^{(k)})\}$$

$$\le c\,(\log n)\{d\, B(S^{(h)}) - d\, B(S^{(k)})\} .$$

As $n \to \infty$, $\limsup \frac{\beta_2(n)}{\log n} \le c\,\{\beta - d\, B(S^{(k)})\}$ by (3.3.2). Therefore, by (3.3.6) and (3.3.8),

$$\limsup_{n\to\infty} \frac{\beta(n)}{\log n} \le d\, B(S^{(k)}) + c\,\{\beta - d\, B(S^{(k)})\} .$$

Since this is true for all k,

$$\overline{\delta}\, B(S) = \limsup_{n\to\infty} \frac{\beta(n)}{\log n} \le \lim_{k\to\infty} d\, B(S^{(k)})$$

$$+ c\,\{\beta - \lim_{k\to\infty} d\, B(S^{(k)})\}$$

$$\le \beta \quad \text{by (3.3.8).}$$

Thus Theorem 3.3 is proven, and this theorem, together with (3.1.6) gives (3.1.2), which proves Theorem 3.1.

In light of the Davenport-Erdos Theorem, it might be conjectured that if $\overline{\delta}\, S > 0$ then $\lim_{x\to\infty} \frac{f(x)}{x} = \infty$. Erdös, Sárközi, and Szemerédi [9] proved that while $\limsup_{x\to\infty} \frac{f(x)}{x} = \infty$ for any such sequence S, for each $0 < c \le 1$, there is a sequence S', with $\overline{\delta}\, S' = c$, for which $\frac{f(x)}{x}$ increases slowly.

<u>Theorem 3.4</u> Suppose S is an infinite sequence with

(3.4.1) $$\bar{\delta} S = \limsup_{x\to\infty} \frac{1}{\log x} \sum_{s_i \le x} \frac{1}{s_i} = c_1 > 0 .$$

Then there is a $c_2 = c_2(c_1)$ so that

(3.4.2) $$f(x) > x \exp\{c_2 \sqrt{\log\log x}\ \log\log\log x\}$$

for infinitely many x. However, there is a sequence S' satisfying (3.4.1) for which

(3.4.3) $$f(x) < x \exp\{c_3 \sqrt{\log\log x}\ \log\log\log x\}$$

for all x, where $c_3 = c_3(c_1)$.

<u>Proof</u>. Suppose (3.4.1) holds. Then for infinitely many N, $\sum_{s_i \le N} \frac{1}{s_i} > \frac{c_1}{2} \log N$. (3.3.2) is a consequence of the following theorem.

<u>Theorem 3.5</u> Let $t_1 < \dots < t_k \le N$ be a sequence of positive integers satisfying

(3.5.1) $$\sum_{t_i \le N} \frac{1}{t_i} > c_4 \log N .$$

Then there is a $c_5 = c_5(c_4)$ so that

(3.5.2) $$\sum_{t_i \le N}{}_1 \frac{1}{t_i} > \frac{1}{2} c_4 \log N$$

for sufficiently large N, where in $\sum_1$, the summation is over all

t's with at least $\exp\{c_5 \sqrt{\log\log N} \log\log\log N\}$ divisors among the t's.

For each N which satisfies (3.5.2), there is an $M = M(N) < N$ for which

$$\sum_{t_i \le M}{}_1 \, 1 > \frac{c_4}{4} M$$

where M tends to infinity with N. Hence

$$f(M) \ge \left(\sum_{t_i \le M}{}_1 \, 1 \right) \exp\{c_5 \sqrt{\log\log N} \log\log\log N\}$$

$$> M \exp\{\frac{c_5}{2} \sqrt{\log\log M} \log\log\log M\}$$

and Theorem 3.4 follows from Theorem 3.5.

Proof of Theorem 3.5 Suppose the theorem is false. Then for arbitrarily large N, there exists a sequence $t_1 < \ldots < t_k \le N$ which satisfies (3.5.1) with, for every c_5, a subsequence $u_1 < \ldots < u_r \le N$ in which each u_i has fewer than $\exp\{c_5 \sqrt{\log\log N} \log\log\log N\}$ divisors among the u's and for which

(3.5.3) $$\sum_{u_i \le N} \frac{1}{u_i} > \frac{1}{2} c_4 \log N \quad .$$

The argument used in Behrend's Theorem shows that there is a V and a subsequence $u_{i_1} < \ldots < u_{i_p}$, with $u_{i_j} = V^2 q_j$ for $j = 1, \ldots, p$, where q_j is square-free, for which

$$\sum_{j=1}^{p} \frac{1}{q_j} > \frac{c_4}{4} \log N \tag{3.5.4}$$

Let $d_5(n)$ be the number of q's dividing n. By our hypothesis we have, for $1 \le r \le p$,

$$d_5(q_r) < \exp \{c_5 \sqrt{\operatorname{loglog} N} \operatorname{logloglog} N\} . \tag{3.5.5}$$

For $N > N_0$, (3.5.4) gives

$$\sum_{m=1}^{N} d_5(m) = \sum_{j=1}^{p} \left[\frac{N}{q_j} \right] \ge N \sum_{j=1}^{p} \left(\frac{1}{q_j} - 1 \right) \tag{3.5.6}$$

$$> \frac{c_4}{5} N \log N .$$

Since each q_j is square-free, $d_5(m) \le 2^{\omega(m)}$. Thus from (3.5.6),

$$\sum_{\substack{m=1 \\ \omega(m) > \operatorname{loglog} N}}^{N} d_5(m) > \frac{c_4}{5} N \log N - N\, 2^{\operatorname{loglog} N} \tag{3.5.7}$$

$$> \frac{c_4}{10} N \log N .$$

However $\displaystyle\sum_{m=1}^{N} d(m) < 2 N \log N$.

Hence by (3.5.7) there is an $m \le N$ with $\omega(m) > \operatorname{loglog} N$ for which

$$d_5(m) \ge \frac{c_4}{20} d(m) \ge \frac{c_4}{20} 2^{\omega(m)} . \tag{3.5.8}$$

We complete the proof by using Theorem 1.10 to obtain a contradiction to (3.5.5).

To apply Theorem 1.10 in the proof of Theorem 3.5, we let S be the set of prime factors of the m of (3.5.8). Since the q's are square-free, we can assume that m is square-free, so $|S| = n = \omega(m)$. Let B_i be the set of prime factors of q_i. Then by (3.5.8), there are more than $\frac{c_4}{20} 2^n$ B's. By Theorem 1.10, one of the B's contains at least $\exp\{c_7 \sqrt{n} \log n\}$ other B's for $n > n_0(c_4)$, where $c_7 = c_7(c_4)$. That is, there is a q dividing m for which

$$d_5(q) > \exp\{c_7 \sqrt{\log\log N} \log\log\log N\}$$

which contradicts (3.5.5) for sufficiently small c_5.

Thus (3.4.2) is proved. As the proof of (3.4.3) requires the use of probabilistic methods, we shall not present the argument here. The result indicates, however, that (3.4.2) is close to being best possible.

BIBLIOGRAPHY

[1] I. Anderson, "On primitive sequences", J. Lond. Math. Soc., Vol. 42 (1967), 137-148.

[2] F. Behrend, "On sequences of integers not divisible one by another", J. Lond. Math. Soc., Vol. 10 (1935), 42-44.

[3] A. Besicovitch, "On the density of certain sequences of integers", Math. Annln., Vol. 110 (1934), 336-341.

[4] H. Davenport and P. Erdös, "On sequences of positive integers", Acta Arith., Vol. 2 (1937), 147-151.

[5] P. Erdös, "Integers with exactly k prime factors", Ann. Math. II, Vol. 49 (1948), 53-66.

[6] P. Erdös, "Note on sequences of integers no one of which is divisible by any other", J. Lond. Math. Soc., Vol. 10 (1935) 126-128.

[7] P. Erdös, A. Sárközi, and E. Szemerédi, "On an extremal problem concerning primitive sequences", J. Lond. Math. Soc., Vol. 42 (1967), 484-488.

[8] P. Erdös, A. Sárközi, and E. Szemerédi, "On a theorem of Behrend", Austral. Math. Soc. J., Vol. 7 (1967), 9-16.

[9] P. Erdös, A. Sárközi, and E. Szemerédi, "On the divisibility properties of sequences of integers I", Acta Arith., Vol. 11 (1966), 411-418.

[10] H. Halberstam and K. Roth, Sequences, Ch. V (Oxford University Press, 1966).

[11] D. Kleitman, "On subsets contained in a family of non-commensurable subsets of a finite set", J. Comb. Th., Vol. 7 (1969), 181-183.

[12] S. Pillai, "On numbers which are not multiples of any others in the set", Proc. Ind. Acad. Sci. A, Vol. 10 (1939), 392-394.

[13] E. Sperner, "Ein Satz über Untermengen einer endlichen Menge", Math. Z., Vol. 27 (1928), 544-548.